Cytochrome P450 enzymes involved in xanthone biosynthesis in *Hypericum* species

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Islam Ahmed Mohamed El-Awaad
aus Gharbeya / Ägypten

1. Referent: Professor Dr. Ludger Beerhues
2. Referentin: Professorin Dr. Ute Wittstock
3. Referentin: Professorin Dr. Anett Schallmey
eingereicht am: 09.11.2015
mündliche Prüfung (Disputation) am: 05.04.2016

Druckjahr 2016

„Gedruckt mit Unterstützung des Deutschen Akademischen Austauschdienstes“

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

El-Awaad, I., Bocola, M., Beuerle, T., Liu, B. & Beerhues, L. Bifunctional CYP81AA proteins catalyse identical hydroxylations but alternative regioselective phenol couplings in plant xanthone biosynthesis. Nature Communications 7, 11472 (2016).

Fiesel, T., Gaid, M., Müller, A., Bartels, J., **El-Awaad, I.**, Beuerle, T., Ernst, L., Behrends, S. & Beerhues, L. Molecular cloning and characterization of a xanthone prenyltransferase from *Hypericum calycinum* cell cultures. Molecules 20, 15616-15630 (2015).

Tagungsbeiträge

El-Awaad, I., Bocola, M., Liu, B. & Beerhues, L. Regioselective C–O phenol couplings catalyzed by xanthone biosynthetic CYPs. (Poster) Trends in Natural Products Research 2014 Young Scientists Meeting, Olomouc, Czech Republic (2014).

El-Awaad, I., Liu, B. & Beerhues L.: Cytochrome P450 enzymes involved in xanthone Biosynthesis. (Lecture) 5th Annual *Hypericum* Meeting, Heidelberg (2012).

El-Awaad, I. & Beerhues L.: cDNA cloning and expression of cytochrome P450 enzymes catalyzing regiospecific C–O couplings in xanthone biosynthesis. (Lecture) 7. Tagung der Sektion ‘Pflanzliche Naturstoffe’ der Deutschen Botanischen Gesellschaft, Hildesheim (2012).

Acknowledgments

All praise to almighty **Allah** for all the blessings in my life.

I am extremely grateful to my supervisor **Prof. Dr. Ludger Beerhues** for giving me the opportunity to join his lab and for suggesting the research topic of this thesis. His continuous guidance and support throughout the phases of this project were essential for its success. I have learnt a lot from him both on the professional and personal sides. His decent character, calm and confident way of facing various situations showed me an example of how one should behave in a scientific environment. During our many discussions, his comments were always constructive. He drew my attention to any mistakes and corrected them in a smooth and friendly manner.

I am very thankful to **Prof. Dr. Ute Wittstock** and **Prof. Dr. Anett Schallmey** for accepting to referee this work.

I extend my deep gratitude to **Dr. Marco Bocola** for performing the homology modeling and product docking works, answering my many questions with patience and for the helpful discussions in designing the successive stages of mutations. I am also thankful to **Dr. Till Beuerle** for guiding me through the chemical synthesis of standards and for his great help with mass spectrometry. I also appreciate the time and efforts of **Prof. Dr. Ludger Ernst** in confirming the identity of the synthesized compounds by NMR. I am thankful to **Dr. Benye Liu** for constructing the subtracted library, on which a major part of this work is based. I appreciate his kindness and welcoming friendly personality. I am also thankful to **Dr. Rainer Lindigkeit** for the technical support and valuable discussions.

I was extremely lucky to work with **Dr. Andreas Müller** at the beginning of my PhD. I appreciate his team spirit, sharing ideas, tips and even mistakes. I still miss our scientific discussions, which were very constructive and opened new gates in my way of thinking and handling problems.

Getting to know **Dr. Mohammed Khalil** was one of the greatest benefits of my stay in Germany. His unique personality spread a friendly atmosphere around the lab. He was always there whenever anyone needed help. I have benefited a lot from his advice and experience. I would like to thank him for the fruitful discussions in various scientific topics.

I was happy to find familiar faces when I first joined the lab. The presence of my former colleagues **Dr. Mariam Gaid** and **Dr. Iman Abdelrahman** accelerated my adaptation inside and outside the lab. I am very thankful to their help and advice.

I would like to thank **Mina Noshy** and **Mohamed Nagia** for the nice time we had together. Their dynamic characters and sense of humor spread fun in the stringent scientific atmosphere.

I also thank **Dr. Marion Wiggermann** and **Dennis Reckwell** for the friendly talks and nice memories we had together especially during our trip to the Czech Republic.

I sincerely appreciate the efforts of **Ines Rahaus**, **Doris Glindemann** and **Bettina Böttner**, which make everyone's job much easier.

Great appreciation and gratitude to all former and present colleagues from the Beerhues group **Dr. Debabrata Sircar**, **Dr. Sahar Abdelaziz**, **Dr. Ines Belhadj**, **Tobias Fiesel**, **Ebtisam Ali**, **Rabeia Ali** and **Marco Grull** for the friendly working atmosphere and their continuous encouragement. Extended thanks to the former and current members of the Wittstock group **Dr. Einar Stauber**, **Dr. Luise Cramer**, **Dr. Maike van Ohlen**, **Dr. Frauke Gumz**, **Nargis Elgahme**, **Malte Büttner**, **Friederike Dörr**, **Kathrin Meier**, **Anita Backenköhler**, **Eline Biedermann**, **Anna-Maria Herfurth** and **Elena Kurzbach** for the nice atmosphere and so many birthday cakes.

I am grateful to the German Academic Exchange Service (**DAAD**) and the Egyptian Ministry of Higher Education (**MOHE**) for granting me the German-Egyptian Research Long-Term Scholarship (**GERLS**) to support my stay in Germany.

I sincerely appreciate the efforts of my wife **Mai**. Thank you for your continuous encouragement, everlasting moral support and understanding. You paved my way and provided an excellent atmosphere so that I can totally concentrate on my work. Thank you for being in my life.

To **Omar** and **Nourin**, when I look at your faces smiling, listen to your voices laughing and hold you in my arms, all tiredness and stress of a hard working day just vanishes. You are and will always be the light of my life.

And last but not least, all my heartily thanks, grand indebtedness and enthusiastic appreciation to my **Parents** and brother **Ehab** for the unfailing love, spiritual support, valuable advice and constant prayers.

Islam El-Awaad

Braunschweig, Germany

To my Parents,

To Mai, Omar and Nourin

Table of Contents

Abbreviations

°C	Degree Celsius
µg	Microgram
µl	Microliter
µM	Micromolar
2,4-D	2,4-Dichlorophenoxyacetic acid
APS	Ammonium peroxydisulfate
ATP	Adenosine triphosphate
B3′H	Benzophenone 3′-hydroxylase
Bam	*Bacillus amyloli*
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BPS	Benzophenone synthase
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
cds	Coding DNA sequence
CoA	Coenzyme A
Contig	A set of overlapping DNA sequences
Cpd	Compound
CPR	NADPH-dependent Cytochrome P450 reductase
cps	Counts per second
Cq	Quantification cycle
CYP	Cytochrome P450 enzyme
CYP	Cytochrome P450 gene
Cyt c	Cytochrome c
DAD	Diode array detector
dATP	Deoxyadenosine triphosphate
DEPC	Diethylpyrocarbonate
dH_2O	Deionized water
diHB	Dihydroxybenzophenone
diHX	Dihydroxyxanthone
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
DTT	1,4-Dithiothreitol
Eco	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EPI	Enhanced product ion
EST	Expessed sequence tag
FAD	Flavin adenine dinucleotide
FdR	Ferredoxin reductase
Fdx	Iron–sulfur-cluster
FMN	Flavin mononucleotide
g	Gravitational force

g	Gram
GSP	Gene specific primer
h	Hour
HB	Hydroxybenzophenone
Hin	*Haemophilus influenzae*
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUPAC	International union of pure and applied chemistry
kDa	Kilodalton
Km	Michaelis-Menten constant
l	Liter
LB	Luria broth
LS	Linsmaier and Skoog
M	Molar
MD	Molecular dynamics
MCS	Multiple cloning site
MIC	Minimum inhibitory concentration
min	Minute
ML	Maximum likelihood
ml	Milliliter
mM	Millimolar
MPGR	Medicinal plants genomics resource
mRNA	Messenger RNA
MS	Mass spectroscopy
NAA	1-Naphthaleneacetic acid
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
ng	Nanogram
NGS	Next generation sequencing
nm	Nanometer
OD	Optical density
ORF	Open reading frame
P450	Cytochrome P450 enzyme
Pac	*Pseudomonas alcaligenes*
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG 4000	Polyethylene glycol 4000
pentaHB	Pentahydroxybenzophenone
pI	Isoelectric point
PKS	Polyketide synthase
PT	Prenyltransferase
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
rpm	Revolution per minute
RT	Reverse transcription
RT-PCR	Reverse transcription polymerase chain reaction

s	Second
s.d.	Standard deviation
s.e.m.	Standard error of the mean
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulfate
SOC	Super optimal broth with catabolite repression
Spe	*Sphaerotilus sp.*
SRS	Substrate recognition site
SSH	Suppression subtractive hybridization
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
TdT	Terminal deoxynucleotidyl transferase
TEMED	N,N,N′,N′-tetramethylethylenediamine
tertaHX	Tetrahydroxyxanthone
tetraHB	Tetrahydroxybenzophenone
T_m	Melting temperature (primer)
triHB	Trihydroxybenzophenone
triHX	Trihydroxyxanthone
Tris	Tris(hydroxymethyl)aminomethane
TXS	Trihydroxyxanthone synthase
UTR	Untranslated region
UV	Ultraviolet
X6H	Xanthone 6-hydroxylase
X-GAL	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside
λ_{max}	Maximum absorption wavelength

Amino acids

Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparine	N	Asn	Methionine	M	Met
Aspartic	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamic	E	Glu	Serine	S	Ser
Glutamine	Q	Gln	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

Nucleotides

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

I. Introduction

1. Cytochrome P450 enzymes

1.1. Nature

Cytochrome P450 (P450 or CYP) enzymes are present in almost all life forms including bacteria, lower eukaryotes, plants and animals and have even been found in viruses (Danielson, 2002; Lamb et al., 2009). Their active site contains an iron-heme prosthetic group, where the nitrogen atoms of the heme provide the first four ligands while a cysteine thiolate forms the fifth ligand of the heme iron. This configuration is responsible for the characteristic absorption maximum at 450 nm upon binding of carbon monoxide to the ferrous heme iron, and hence the name 'Cytochrome P450' or the Pigment absorbing at 450 nm (Omura and Sato, 1964). The enzymes can activate molecular oxygen (O_2) and, generally, insert one atom in the substrate, while the second atom is reduced to water according to the general equation: $RH + O_2 + 2e^- + 2H^+ \longrightarrow ROH + H_2O$, where RH is the substrate and ROH is the hydroxylated product (Bak et al., 2011). CYPs catalyze a wide array of reactions in primary (general) and secondary (specialized) metabolism and play a major role in xenobiotic detoxification. Common CYP-catalyzed reactions include C-hydroxylation, dealkylation and epoxidation. However, some CYPs have been reported to catalyze more complex, less common reactions such as dimerization and coupling as well as ring formation, contraction, and expansion (Guengerich, 2001).

1.2. Nomenclature

Unlike animals and bacteria, plants are characterized by the presence of large numbers of CYP genes representing about 1% of their genomes (**Table I-1**). Currently, there are about 21,000 annotated CYPs, more than 7400 being found in the plant kingdom (Nelson, 2009). It is expected that the number will further increase as a result of the blooming genome and transcriptome sequencing projects supported by the recent developments in next generation sequencing (NGS) techniques.

Table I-1. Examples of CYP gene and pseudogene counts in selected organisms.

Organism	CYP genes count	CYP pseudogenes count
Homo sapiens	57	47
Mus musculus	103	82
Arabidopsis thaliana	244	28
Ricinus communis	214	47
Glycine max	313	195
Populus tichocarpa	310	203
Vitis vinifera	315	201
Oryza sativa	323	135
Solanum lycopersicum	366	93

A common nomenclature system based on the level of amino acid sequence identity and phylogenetic association has been devised to categorize CYP genes from all organisms. The name consists of the root symbol 'CYP', representing 'cytochrome P450', followed by an Arabic number

denoting the family, one or two letters designating the subfamily, and an Arabic numeral representing the individual gene within the subfamily (Nelson et al., 1996). For example, cinnamate 4-hydroxylase from *A. thaliana* was given the name CYP73A5 (Mizutani et al., 1997). This denotes that this protein belongs to family 73, subfamily A, and is the peptide number 5 within this subfamily. As a general rule, CYPs sharing at least 40% amino acids sequence identity are grouped within the same family, while those sharing at least 55% identity are grouped within the same subfamily (Nelson, 2006). Since CYPs are found in all domains of life, blocks of numbers have been reserved for the classification of new CYP families according to their taxonomic divisions. The family numbers has now increased from two to four digits to accommodate the expanding number of newly discovered CYPs (**Table I-2**) (Danielson, 2002).

Table I-2. CYP families ranges for the major taxonomic divisions.

Taxonomic division	CYP families range		
Animals	CYP1-49	CYP301-499	CYP3001-4999
Lower eukaryotes	CYP51-69	CYP501-699	CYP5001-6999
Plants	CYP71-99	CYP701-999	CYP7001-9999
Bacteria	CYP101-299	CYP1001-2999	

1.3. Structure, conserved domains and substrate recognition

CYPs have variable substrate specificities and share low primary amino acid sequence identities that may be less than 20% between members of different families (Graham and Peterson, 1999). However, the increasing number of solved crystal structures show that most CYPs have common features. While prokaryotic CYPs are soluble proteins, eukaryotic CYPs are associated with microsomal membranes by a hydrophobic N-terminal membrane anchor that is separated from the globular part of the protein by a cluster of basic residues and a proline-rich motif (usually PPxP), which acts as a hinge between both parts (**Figure I-1**) (Werck-Reichhart and Feyereisen, 2000).
The common conserved secondary structures of CYPs include twelve α helices (αA-L) and five β sheets (β1-5), which could be identified in all CYP structures and make up the common CYP-fold (**Figure I-2**) (Sirim et al., 2010). The heme prosthetic group is the catalytic center of the enzymes and is found deeply buried inside all CYP proteins at the bottom of a large, internal binding cavity (**Figure I-3**) (Seifert and Pleiss, 2009). The core part of the protein surrounding the heme shows highest structural conservation reflecting a common mechanism of oxygen activation and electron or proton transfer. This core region includes:

- The heme-binding loop which is located just before the L-helix and contains the absolutely conserved cysteine in the signature sequence (FxxGxRxCxG). The underlined cysteine serves as the proximal fifth ligand to the iron of the heme.
- The conserved ExxR motif which is located in the K-helix on the heme proximal side. It is important for the stabilization of the meander loop and, in turn, the CYP tertiary structure.
- The I-helix which is located distal to the heme with its central part directly adjacent to the iron. It contains the consensus sequence A/GGxD/ETT/S. The underlined conserved threonine plays an important role in proton transfer during catalysis.

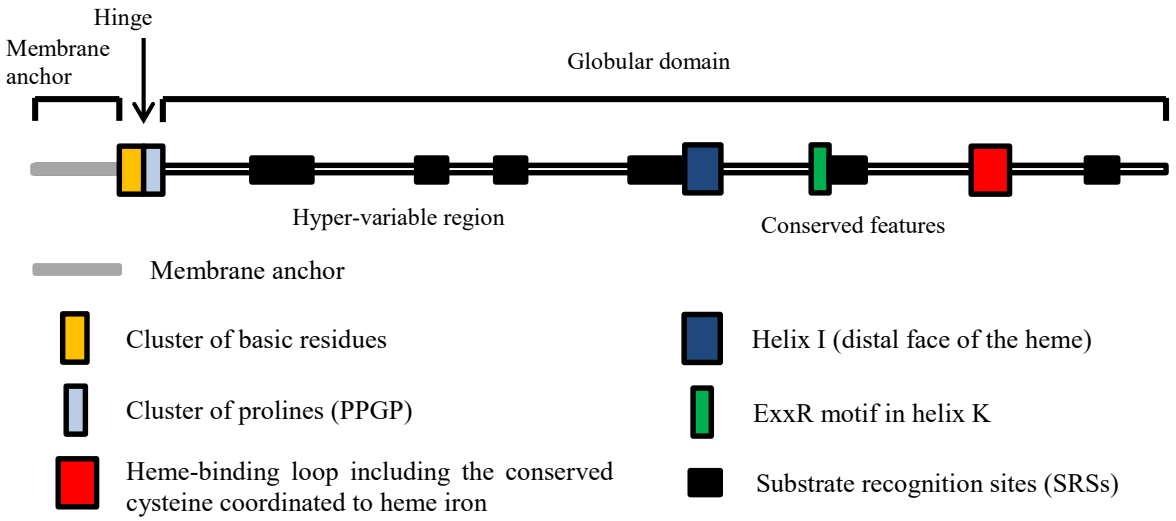


Figure I-1. Primary structure of endoplasmic reticulum-bound CYPs (class II); adapted from Werck-Reichhart and Feyereisen (2000).

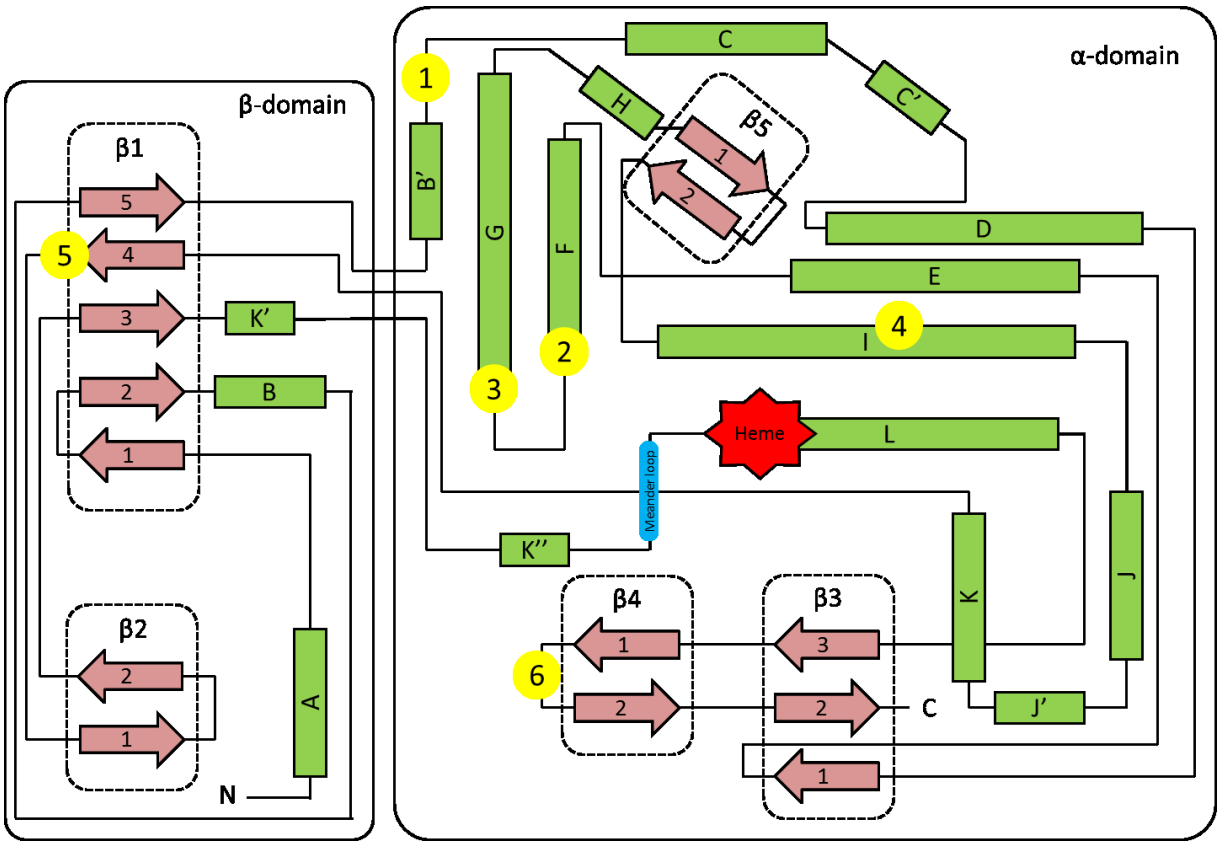


Figure I-2. A topological illustration of CYPs showing the main structural features, the substrate recognition sites (SRSs) being highlighted in yellow; adapted from Werck-Reichhart and Feyereisen (2000) and Sirim et al. (2010).

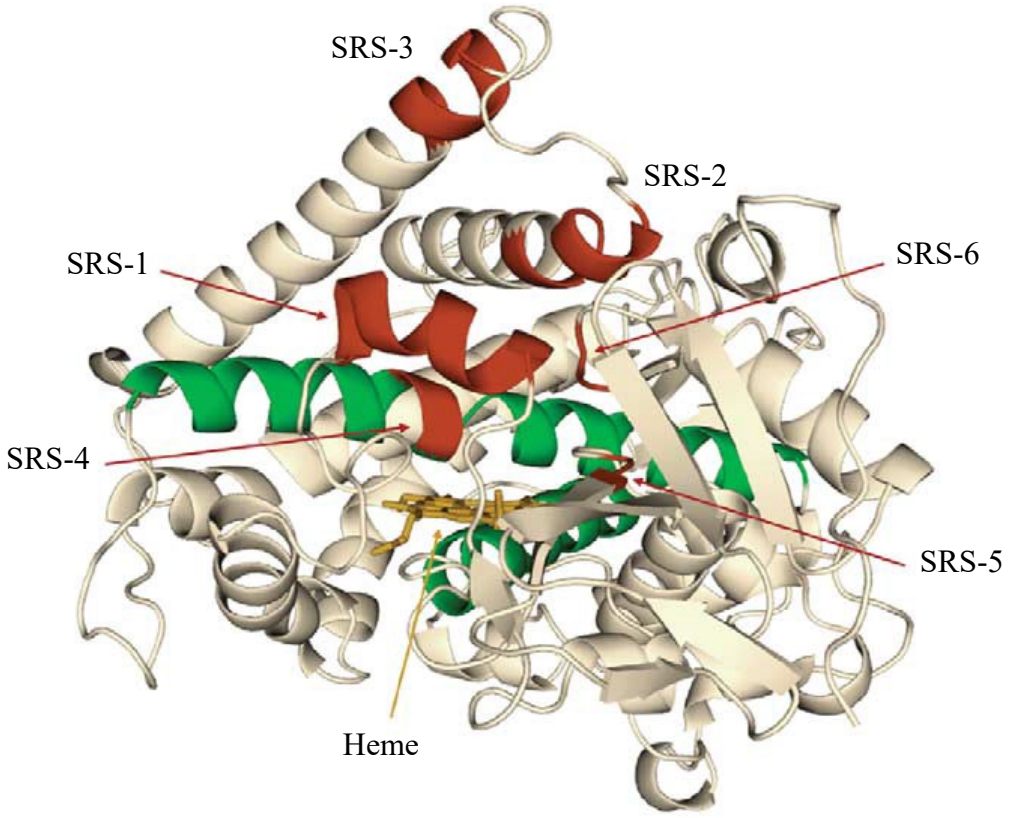


Figure I-3. Cartoon illustration of the tertiary structure of CYPs. The heme is colored orange, the substrate recognition sites (SRSs 1–6) are marked red, and the I and L helices are shown in green; adapted from Urlacher and Eiben (2006).

On the other hand, substrate binding and recognition are mediated by variable regions known as the substrate-recognition sites (SRSs). The SRSs were first defined by Gotoh (1992) upon alignment of the mammalian CYP2 family with the bacterial CYP101A (P450cam), which was the first CYP crystal structure with the bound camphor substrate to be solved (Poulos et al., 1987). Regions designated to form the SRSs include (**Figure I-2**, **I-3**):

- SRS-1: the region between the B and C-helices (including the B'-helix).
- SRS-2: the C-terminal end of the F-helix.
- SRS-3: part of the F-G loop and N-terminal end of the G-helix.
- SRS-4: the N-terminal part of the I-helix.
- SRS-5: the loop between the K helix and strand 4 of β-sheet 1.
- SRS-6: the turn in β-sheet 4.

The residues of the first three SRSs surround the substrate-access channel and are thought to be responsible for the substrate specificity, while the residues in the latter three SRSs are much closer to the heme catalytic center and most likely present in direct contact with the substrate during catalysis. Therefore, they play an important role in controlling the regio- and stereoselectivities of the CYPs.

An updated map of the assigned SRSs based on docking of 868 different substrates to mammalian CYPs was recently introduced (Zawaira et al., 2011). The Zawaira map differs from the Gotoh map in the following aspects:

- The length of SRS-1 increased from 28 residues to 49 residues.
- Two additional SRS-1'a and SRS-1'b regions located N-terminally to SRS-1 were identified.
- SRS-2 and SRS-3 were merged into SRS (2, 3) with a total length of 60 residues instead of 19 in the Gotoh map.
- The lengths of SRSs 4, 5, and 6 increased by a few residues.

Overall, about 33% of the residues on a CYP are assumed to be within the SRSs in the new map compared to only 16% in the Gotoh map (Zawaira et al., 2011; Mustafa et al., 2014).

1.4. Catalytic cycle

In the resting state, the iron is hexacoordinated by the four nitrogen atoms of the heme with the sulphur of the conserved cysteine at the proximal and a water molecule at the distal sides. The iron atom exists in the ferric low-spin state (Ogliaro et al., 2002; Lamb and Waterman, 2013).
The catalytic cycle of CYPs involves the following steps (**Figure I-4**):

1. **Substrate binding:** When the substrate binds to the protein, the distal water molecules is removed. Concomitantly, the ferric iron undergoes transition from the low-spin to the high-spin state with increasing its redox potential.
2. **First one-electron reduction:** The redox partner (for example, the NADPH-dependent cytochrome P450 reductase in eukaryotes) transfers an electron to the CYP, resulting in the formation of a ferrous pentacoordinated heme.
3. **Dioxygen binding:** Molecular oxygen binds to the heme iron resulting in the formation of the ferric superoxo species.
4. **Second one-electron reduction:** The redox partner supplies a second electron to the CYP resulting in the formation of a ferric peroxo species.
5. **Protonation:** A ferric hydroperoxo species (also known as Cpd 0) is generated when the previously mentioned ferric peroxo species receives a proton from its environment.
6. **Second protonation and heterolytic scission of the O–O bond:** A second protonation together with heterolytic cleavage of the O–O bond leads to the release of water and the generation of the reactive iron(IV)oxo (or ferryl-oxo) π-radical species known as Cpd I (Rittle and Green, 2010).
7. **Reaction with the substrate:** The highly reactive Cpd I abstracts a hydrogen atom from the substrate to form a substrate radical and a ferryl-hydroxo species known as Cpd II. In the classical hydroxylation mechanism, oxygen rebound occurs leading to the formation of the hydroxylated product while the heme iron returns to the ferric state.
8. **Release of the product:** When the hydroxylated product is released, a water molecule occupies back the sixth coordinate of the heme and the enzyme returns to its resting state ready for another cycle.

Binding of the reduced form of the enzyme to carbon monoxide gives rise to the characteristic Soret absorption peak at 450 nm due to the properties of the cysteine thiolate group (**Figure I-4**, step 9). Another way for heme activation is the peroxide shunt (**Figure I-4**, step 10), where the interaction with peroxides and hypochlorites as single-oxygen donors can lead to the formation of Cpd 0 directly without going through steps 2–5 of the cycle. This phenomenon was recently employed to characterize the nature of Cpd I (Rittle and Green, 2010).

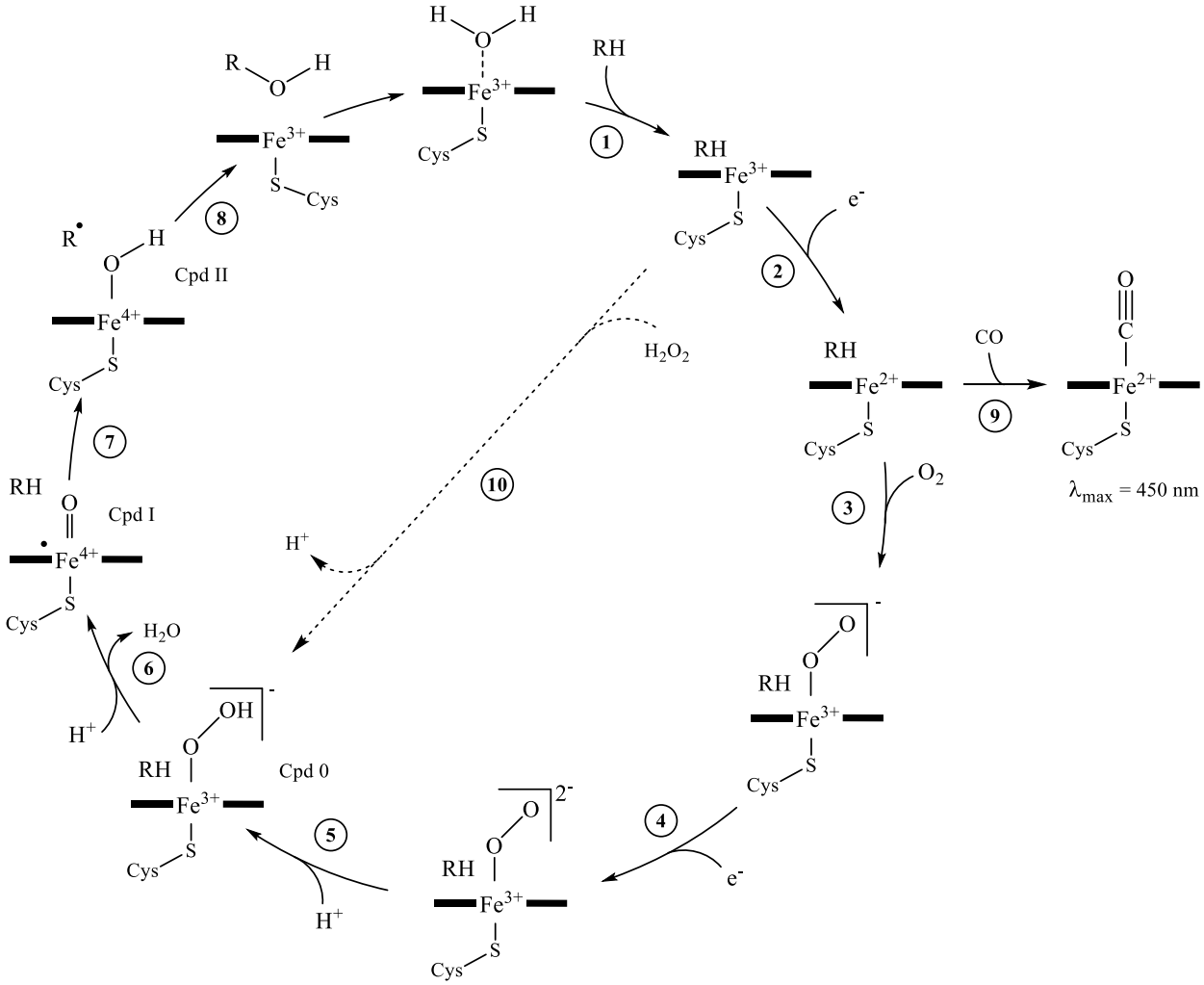


Figure I-4. Catalytic cycle of CYP-catalyzed hydroxylation. RH is the substrate, ROH is the hydroxylated product.

1.5. CYP redox partners

As inferred from the catalytic cycle, the vast majority of CYPs require supply with two electrons per cycle to accomplish substrate hydroxylation. Those electrons should be delivered from NAD(P)H cofactor one at a time by the aid of a redox partner. When CYPs were first discovered in human adrenal gland and liver, two main classes of redox partners were identified. In adrenal mitochondrial CYPs, adrenodoxin and adrenodoxin reductase were involved (Omura et al., 1966), whereas in liver microsomal systems, NADPH dependent oxidoreductase containing FAD and FMN was involved (Lu et al., 1969). The first bacterial CYP discovered was camphor hydroxylase

(CYP101 or P450cam) and was identified as a soluble protein (Katagiri et al., 1968). It receives its electrons from a putidaredoxin/putidaredoxin reductase system analogous to the mitochondrial one. Later on, CYP102 (P450BM3) was discovered as a fusion protein consisting of both CYP and reductase domains (Narhi and Fulco, 1987; Hannemann et al., 2007).
The above mentioned systems represent the most abundant types of redox partners involved in CYP catalysis. However, other less abundant systems exist in nature as well. Currently, a total of 10 classes of CYP redox partner systems have been identified, as summarized in (**Table I-3**).

Table I-3. Classes of CYPs based on their redox partner systems; adapted from Hannemann et al. (2007).

Class/source	Electron transport chain	Localization/remarks
Class I		
Bacterial	NAD(P)H ▸ [FdR] ▸ [Fdx]* ▸ [P450]	Cytosolic, soluble
Mitochondrial	NADPH ▸ [FdR] ▸ [Fdx] ▸ [P450]	FdR: membrane associated, Fdx: mitochondrial matrix, soluble
Class II		
Bacterial	NADH ▸ [CPR] ▸ [P450]	Cytosolic, soluble
Microsomal A	NADPH ▸ [CPR] ▸ [P450]	Membrane anchored, ER
Microsomal B	NADPH ▸ [CPR] ▸ [Cytb5] ▸ [P450]	Membrane anchored, ER
Microsomal C	NADH ▸ [Cytb5Red] ▸ [Cytb5] ▸ [P450]	Membrane anchored, ER
Class III		
Bacterial	NAD(P)H ▸ [FdR] ▸ [Fldx] ▸ [P450]	Cytosolic, soluble
Class IV		
Bacterial	Pyruvat, CoA ▸ [OFOR] ▸ [Fdx] ▸ [P450]	Cytosolic, soluble
Class V		
Bacterial	NADH ▸ [FdR] ▸ [Fdx–P450]	Cytosolic, soluble
Class VI		
Bacterial	NAD(P)H ▸ [FdR] ▸ [Fldx–P450]	Cytosolic, soluble
Class VII		
Bacterial	NADH ▸ [PFOR–P450]	Cytosolic, soluble
Class VIII		
Bacteria, fungi	NADPH ▸ [CPR–P450]	Cytosolic, soluble
Class IX		
fungi	NADH ▸ [P450]	Cytosolic, soluble
Class X		
Independent in plants/mammals	[P450]	Membrane bound, ER

Abbreviated protein components contain the following redox centers: Fdx (iron–sulfur-cluster); FdR, Ferredoxin reductase (FAD); CPR, cytochrome P450 reductase (FAD, FMN); Cytb5, Cytochrome b5; Cytb5Red, Cytochrome b5 reductase; Fldx, Flavodoxin (FMN); OFOR, 2-oxoacid:ferredoxin oxidoreductase (thiamin pyrophosphate, [4Fe–4S] cluster); PFOR, phthatate-family oxygenase reductase (FMN, [2Fe–2S] cluster).
[a] Fdx containing iron–sulfur-cluster of [2Fe–2S], [3Fe–4S], [4Fe–4S], [3Fe–4S]/[4Fe–4S] type.

1.6. Phylogenetic classification of cytochrome P450 enzymes

Initially, a primitive classification scheme divided plant CYPs into A-type and non-A-type. This was based on the fact that in a phylogenetic analysis the A-type CYPs constituted a monophyletic tree, while the non-A-type CYPs were more divergent and did not constitute a consistent group. It was believed that A-type CYPs are mainly devoted to plant secondary metabolism, such as the biosynthesis of flavonoids, alkaloids, cyanogenic glycosides and lignans, while non-A-type CYPs were assumed to be mainly involved in primary metabolism, such as sterols and hormones biosynthesis and metabolism. However, this belief proved to be incorrect as many non-A-type CYPs were found to be involved in plant secondary metabolism (Nelson and Werck-Reichhart, 2011; Hamberger and Bak, 2013).

As mentioned in (I.1.2), CYPs have been named into families and subfamilies. However, the proximity of designated family numbers does not necessarily reflect a close phylogenetic relationship between them. Family numbers are given according to the priority of sequence submission to the nomenclature committee. Therefore, to understand the phylogenetic relationships between different families, plant CYPs have been classified into clans; each clan is given the lowest number of the CYP families included (**Table I-4**). So far, 11 plant CYP clans have been identified, each of them being derived from a single ancestor. Seven clans are represented by one family each and are, therefore, termed the single-family clans, while the remaining four are represented by several families each and are termed the multi-family clans (**Table I-4**) (Nelson et al., 2004; Bak et al., 2011; Hamberger and Bak, 2013).

Table I-4. Plant CYP clans and the classification of included families; adapted from (Xu et al., 2015).

Clan	Count	Names of families included						
CYP51	1	CYP51						
CYP71	26	CYP71	CYP73	CYP75	CYP76	CYP77	CYP78	CYP79
		CYP80	CYP81	CYP82	CYP83	CYP84	CYP89	CYP92
		CYP93	CYP98	CYP99	CYP701	CYP703	CYP705	CYP706
		CYP712	CYP719	CYP723	CYP726	CYP736		
CYP72	8	CYP72	CYP709	CYP714	CYP715	CYP721	CYP734	CYP735
		CYP749						
CYP74	1	CYP74						
CYP85	15	CYP85	CYP87	CYP88	CYP90	CYP702	CYP707	CYP708
		CYP716	CYP718	CYP720	CYP724	CYP725	CYP728	CYP729
		CYP733						
CYP86	7	CYP86	CYP94	CYP96	CYP704	CYP730	CYP731	CYP732
CYP97	1	CYP97						
CYP710	1	CYP710						
CYP711	1	CYP711						
CYP727	1	CYP727						
CYP746	1	CYP746						

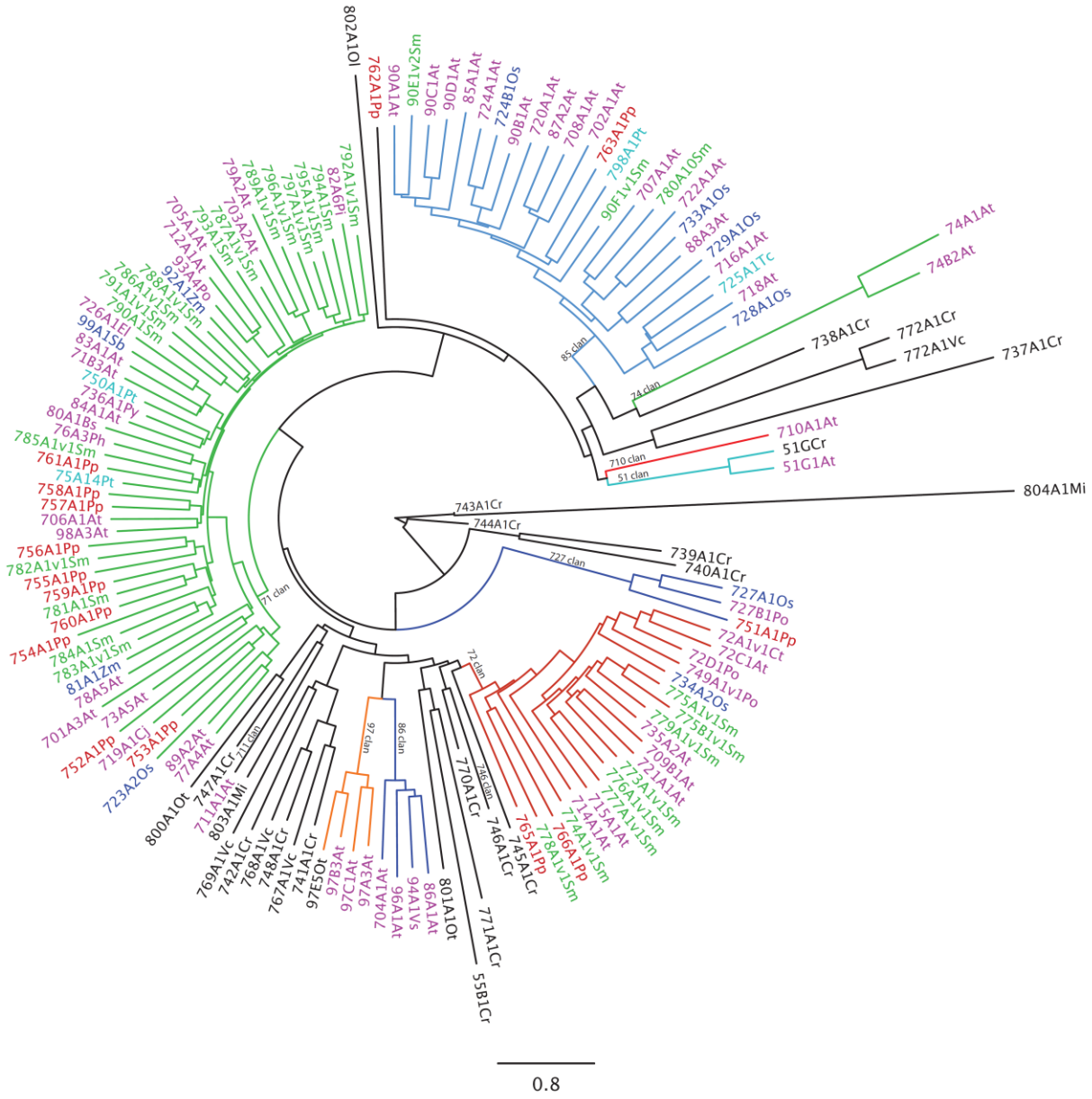


Figure I-5. Phylogenetic analysis of representative members of 127 plant CYP families showing the relationship between the different clans. The analysis was performed by neighbor-joining method. Each clan is labeled and its branches are shown in a different color than the other clans. Algal sequences are shown with black branches and mostly did not fall within the named clans. CYP names are color-coded: magenta, dicots; blue, monocots; cyan, gymnosperms; green, *Selaginella moellendorffii*; red, *Physcomitrella patens*. Two-letter codes are used for species: At, *Arabidopsis thaliana*; Bs, *Berberis stolinifera*; Cj, *Coptis japonica*; Cr, *Chlamydomonas reinhardtii*; Ct, *Catharanthus roseus*; El, *Euphorbia lagascae*; Mi, *Micromonas pusilla*; Ol, *Ostreococcus lucimarinus*; Os, *Oryza sativa*; Ot, *Ostreococcus tauri*; Ph, *Petunia hybrida*; Pi, *Pisum sativum*; Po, *Populus trichocarpa*; Pp, *Physcomitrella patens*; Pt, *Pinus taeda*; Py, *Pyrus communis*; Sb, *Sorghum bicolor*; Sm, *Selaginella moellendorffii*; Tc, *Taxus cuspidata*; Vc, *Volvox carteri*; Vs, *Vicia sativa*; Zm, *Zea mays*; acquired from Nelson and Werck-Reichhart (2011).

Single-family clans generally encode essential enzymes such as those responsible for the production of sterols, lipids, carotenoids and hormones. They occur under purifying selection slowing their duplication and neofuctionalization (Nelson and Werck-Reichhart, 2011). On the

other hand, multi-family clans went through several steps of gene-duplication and neofunctionalization. In the current classification, CYPs formerly classified as A-type exist within the CYP71 clan, while the non-A-Type CYPs are distributed in the other 10 clans. Generally, more than 50 % of higher plant CYPs exist within the CYP71 clan. In spite of being the youngest clan, it underwent several gene duplication events at a quick rate leading to blooms of genes and the emergence of species-specific families. Those were subsequently recruited in secondary metabolism aiding the plants to evolve and adapt to their environment (**Figure I-5**) (Nelson and Werck-Reichhart, 2011).

Although it is difficult to predict CYP substrates from sequence data alone, phylogeny may give a clue about the metabolic pathways in which a certain CYP is involved. For example, members of the CYP85 clan are mainly involved in the metabolism of medium to large isoprenoids such as brassinosteroids, whereas members of the CYP72 clan are involved in the biosynthesis of cytokinins and in the metabolism of a variety of hydrophobic compounds such as gibberellin and brassinosteroids. Finally, CYP71 clan members are mainly recruited in plant specialized metabolism (Nelson et al., 2004; Bak et al., 2011; Nelson and Werck-Reichhart, 2011).

2. The genus *Hypericum*

The genus *Hypericum* comprises about 500 species of herbs, shrubs and less frequently small trees (APGIII, 2009; Nurk et al., 2013). It is therefore by far the largest of the nine genera composing the family Hypericaceae, representing more than 80 % of its species. The Hypericaceae is classified in the clusioid clade of the order Malpighiales (Wurdack and Davis, 2009). The clusioids include five families, namely Bonnetiaceae, Clusiaceae, Calophyllaceae, Podostemaceae and Hypericaceae. Formerly, Hypericaceae family members were treated as a subfamily (Hypericoideae) within the Clusiaceae (Guttiferae). However, after recent phylogenetic analysis of the flowering plant order Malpighiales, they were separated and regarded as distinct families (Stevens, 2007; Crockett and Robson, 2011). Generally, *Hypericum* is distributed worldwide with the major occurrence in Eurasia and Andean South America. It is mainly present in temperate regions as well as highly elevated areas in the tropics (Robson, 2003).

The name *Hypericum* originates from the Greek word *hyper* meaning ‘above’ and *eikon* meaning ‘icon or picture’ because, in the Middle Ages, it was believed that *Hypericum* flowers avert evil spirits and therefore were put above pictures or statues of religious figures. Apparently, this nomenclature reflects the antidepressant activity of the plant because, at that time, it was believed that depression and related conditions are due to evil spirits inhabiting the sick person. As *Hypericum* could alleviate those conditions, people considered it to have ‘the power to ward off evil spirits’ in humans and hence used it to protect their religious figures from demons (Nurk, 2011). *Hypericum* is also commonly known as St. John’s wort as its yellow flowers bloom around St. John’s day on June the 24th (Miller, 1998).

Various *Hypericum* species have been used in traditional medicine in many countries around the globe. This is attributed to the variety of bioactive secondary metabolites synthesized by members of the genus (**Table I-5**) (Butterweck, 2003; Crockett and Robson, 2011; Zhao et al., 2015).

Table I-5. Main classes of secondary metabolites isolated from *Hypericum* species.

Class	Example	Organ
Naphthodianthrones	Hypericin, Pseudohypericin	Aerial parts, in the dark glands
Phloroglucinol derivatives	Hyperforin, Adhyperforin Hyperfirin, Adhyperfirin	Leaves and flowers, in the translucent glands
Benzophenone derivatives	Sampsoniones, Cariphenone Hypelodins	Mainly in roots
Xanthones	Kielcorin, Mangiferin Norathyriol	Mainly in roots
Flavonoids	Rutin, Quercetin Quercitrin, Hyperoside	Aerial parts
Biflavonoids	I3,II8-biapigenin I3′,II8-biapigenin	Flowers
Tannins and procyanidins	Procyanidin, Catechin Epicatechin	Mainly in leaves
Phenylpropanoids	Caffeic acid Chlorogenic acid	
Essential oil	α- and β-Pinene Myrcene, Limonene	Flowers and leaves, in the translucent glands

In the course of this work, two *Hypericum* species have been employed; namely, *Hypericum perforatum* L. and *Hypericum calycinum* L.

2.1. *Hypericum perforatum* L.

Hypericum perforatum L. (common St. John's wort) is a perennial herb distributed mainly in Europe, Asia, North America and northern Africa (Bilia et al., 2002). It is erect, growing up to a height of 100 cm. The leaves are small (1–3 cm long), oblong to elliptic with entire margin and obtuse apex. They contain abundant translucent glands filled with colorless essential oil. The glands span the entire mesophyll which gives the leaves perforated appearance when held against light (**Figure I-6**), and hence the name *perforatum*. The calyx is composed of five sepals, the corolla of five yellow petals, both petals and sepals as well as the leaves are marked with black dots towards their rims. Those dots are known as the dark nodules, filled with reddish brown hypericin pigment that exudates upon rubbing, staining the fingers red (Bilia et al., 2002). *H. perforatum* is the most famous and thoroughly investigated member of the family Hypericaceae (Nahrstedt and Butterweck, 2010). It has been frequently used in traditional medicine because of its wound-healing, anti-inflammatory, diuretic and sedative properties (Di Carlo et al., 2001; Bilia et al., 2002; Butterweck, 2003). The plant is also under intensive investigation for its antiviral, antibacterial and antitumoral effects (Kubin et al., 2005; Medina et al., 2006). The alcoholic extract of the dried flowers or the entire aerial parts is widely prescribed for the treatment of mild to moderate depression (Kasper et al., 2010). *H. perforatum* total sales exceeded 5.5 million dollars

in the United States alone, making it the 26th of the top-selling herbal dietary supplements in the country in 2013 (Lindstrom et al., 2014).

Figure I-6. *H. perforatum* flowers and a leaf held against light showing white and black dots.

Various clinical trials concerning its antidepressant activity have shown that St. John's wort extract is comparable to standard synthetic antidepressants but with fewer side-effects (Di Carlo et al., 2001; Bilia et al., 2002; Schulz, 2006; Linde et al., 2008). Those side effects include photosensitivity when administrated at high doses which is caused by hypericin, pseudohypericin and to lesser extent by hyperforin (Onoue et al., 2011). In addition, St. John's wort extract induces the P450 metabolizing enzymes CYP3A4, CYP2C9 and CYP1A2, leading to pharmacokinetic interactions with drugs metabolized by those enzymes including HIV protease inhibitors, immunosuppressants and coumarin-type anticoagulants (Wang et al., 2001; Henderson et al., 2002).

The mechanism of the antidepressant action of St. John's wort extract is proposed to involve non-selective inhibition of the uptake of neurotransmitters (dopamine, norepinephrine and serotonin) from the synaptic cleft between neurons (Neary and Bu, 1999). Recently, TRPC6 has been detected as hyperforin target (Leuner et al., 2007). Another suggested mechanism was binding to gamma aminobutyric acid (GABA) receptors (Müller et al., 1997). St. John's wort has also been proven to have an inhibitory activity on both monoamine oxidase (MAO) and catechol-O-methyltransferase (COMT) enzymes, leading to an increased level of norepinephrine in the brain (Thiede and Walper, 1994).

The various classes of active constituents in the plant appear to contribute in different levels to its antidepressant activity. It is most likely that a cocktail of those ingredients mediates the pharmacological action through a combination of the aforementioned mechanisms (Di Carlo et al., 2001; Butterweck, 2003; Müller, 2003).

2.2. *Hypericum calycinum* L.

Hypericum calycinum L. (Aaron's beard or creeping St. John's wort) is an evergreen shrub growing to 30 cm height by 1 meter width, usually used as a ground cover. It is distributed mainly in Europe and North America. The leaves are large (up to 10 cm long), oval to oblong, with entire margins and round to obtuse apices. The flowers are 5–7 cm in diameter, with five bright yellow petals and numerous, bushy stamens having reddish anthers. Flowers appear singly or in groups of 2–3 and appear only for a few weeks during the summer (**Figure I-7**). Translucent and dark glands are usually absent from leaves and flowers.

Figure I-7. *Hypericum calycinum* flower showing bright yellow petals and bushy stamens with reddish anthers.

Few studies are available regarding the secondary metabolites produced by *H. calycinum*. Those include phloroglucinol derivatives, cyclohexadienone derivatives, xanthones, and flavonoids (Decosterd et al., 1989; Decosterd et al., 1991; Kirmizibekmez et al., 2009; Win et al., 2012). In addition, it has been reported that tissue cultures of *H. calycinum* can produce adhyperforin and hyperforin under certain conditions (Klingauf et al., 2005). The reported biological activities of the isolated compounds include antioxidant, antifungal, antimalarial, and cell growth inhibitory activities. Moreover, a comparison between the antidepressant activities of several *Hypericum* species extracts suggested that the effect of *H. calycinum* is as potent as that of *H. perforatum* and that it may be used therapeutically in the treatment of depression (Öztürk et al., 1996).

3. Xanthones

Xanthones (9*H*-xanthen-9-one) are a group of secondary metabolites present in a few members of fungi, lichens and higher plant families. The name "xanthone" is derived from the Greek word "xanthos" meaning yellow or blond, referring to their color (Masters and Bräse, 2012). Their basic structure is based on the planar dibenzo-γ-pyron skeleton (**Figure I-8**). However, one of the constituting aromatic rings could be partially or completely saturated giving rise to the di-, tetra-

or hexahydroxanthone skeletons (**Figure I-8**). Because of the symmetric structure of the xanthone nucleus, together with its mixed biosynthetic origin in higher plants, the acetate-derived ring is designated as ring A and is numbered 1–4, while the shikimate-derived ring is designated as ring B and is numbered 5–8 (Bennett and Lee, 1989; Peres and Nagem, 1997). In certain cases, ring B should be given the lowest numbers according to IUPAC nomenclature rules, for example, when ring A is not oxygenated. However, when discussing the biosynthetic aspects of xanthones, the aforementioned numbering and ring designations should be used whenever relevant (Peres and Nagem, 1997). In addition to their chemotaxonomic value, xanthones possess a wide array of interesting pharmacological activities (El-Seedi et al., 2010).

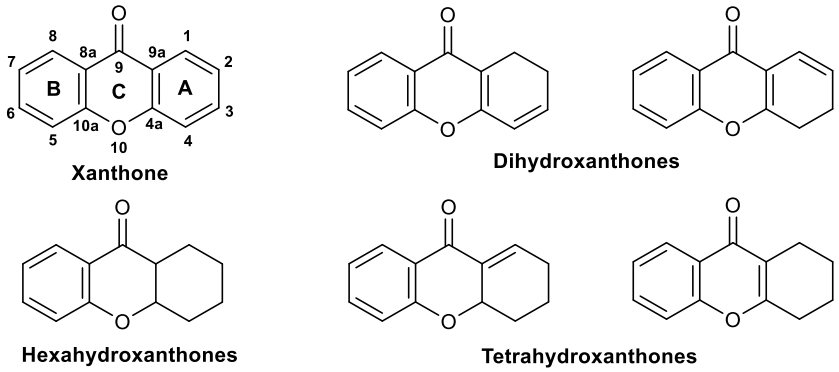


Figure I-8. Basic skeletons of naturally occurring xanthones.

3.1. Classification and occurrence

A review of naturally occurring xanthones in SciFinder® reveals the presence of more than 1500 xanthone-based molecules isolated from natural origin. Those can generally be classified into six major groups: simple xanthones, xanthone glycosides, prenylated xanthones, xanthonolignoids, bis-xanthones and miscellaneous xanthones (Vieira and Kijjoa, 2005).

Occurrence of xanthones in the plant kingdom is restricted to few families, mainly to members of the clusioid clade of the order Malpighiales including the Clusiaceae, Hypericaceae, and Calophyllaceae, in addition to the Gentianaceae, Polygalaceae, Moraceae and Fabaceae (Vieira and Kijjoa, 2005; El-Seedi et al., 2009; El-Seedi et al., 2010).

3.2. Biosynthesis of xanthones

Two distinct pathways are known for xanthone biosynthesis depending on the producing organism. In plants, the xanthone nucleus is synthesized via a mixed acetate-shikimate pathway. Biosynthesis is initiated by benzophenone synthase (BPS), a type III polyketide synthase, which catalyzes the iterative condensation of benzoyl- and 3-hydroxybenzoyl-CoA with three molecules of malonyl-CoA to give 2,4,6-trihydroxybenzophenone and 2,3′,4,6-tetrahydroxybenzophenone (2,3′,4,6-tetraHB), respectively. The former product was found to be converted to 2,3′,4,6-tetraHB by benzophenone 3′-hydroxylase (B3′H) activity (Schmidt and Beerhues, 1997; Liu et al., 2003). The key intermediate 2,3′,4,6-tetraHB undergoes regioselective intramolecular oxidative C–O phenol

coupling reactions either *para* or *ortho* to the 3′-hydroxyl group to give 1,3,7- and 1,3,5-trihydroxyxanthones, respectively, as shown in cell cultures of *Hypericum androsaemum* and *Centaurium erythraea* (**Figure I-9**) (Peters et al., 1997).

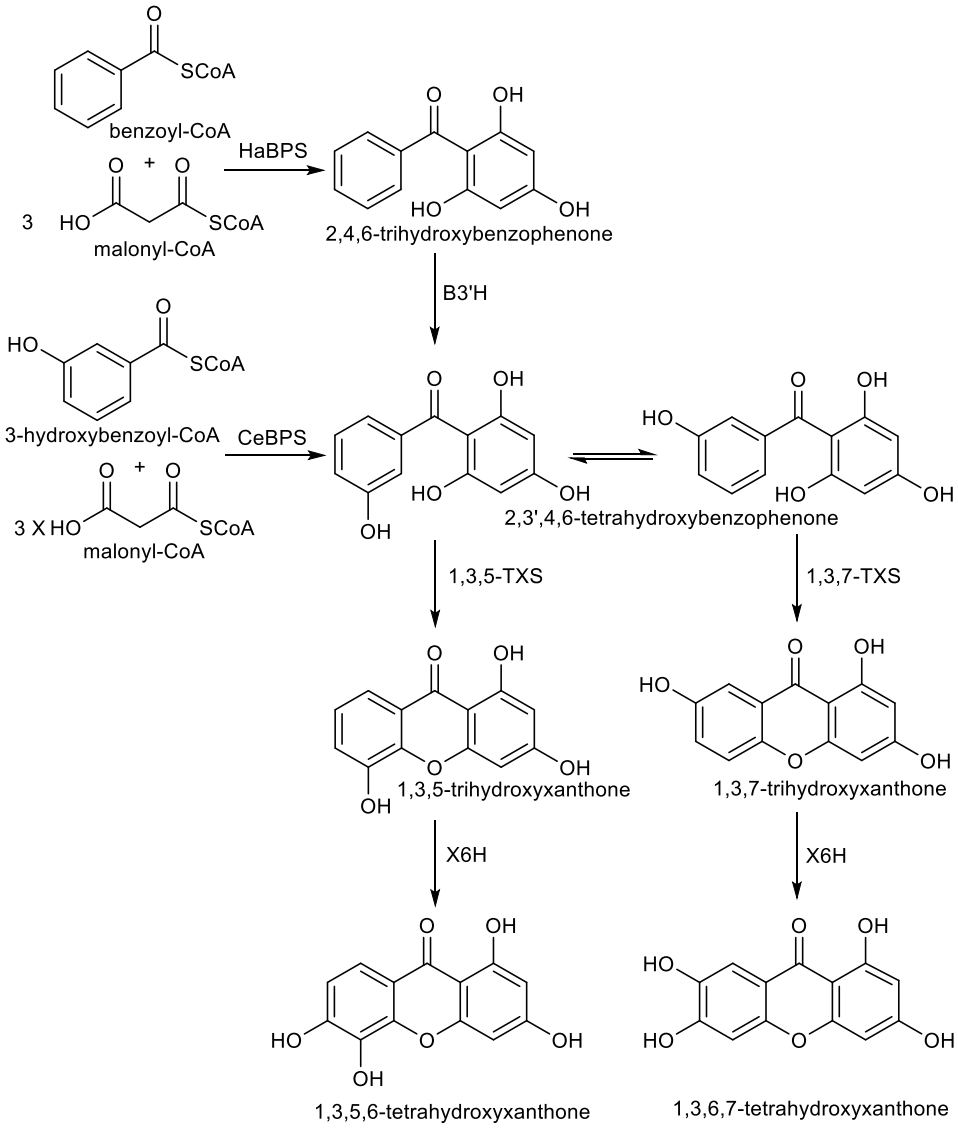


Figure I-9. Biosynthesis of xanthones in plants.

These cyclization reactions are catalyzed by two distinct P450s known as 1,3,7- and 1,3,5-trihydroxyxanthone synthases (TXSs) (Peters et al., 1997). The isomeric products of the cyclization reactions are likely to be the precursors of all plant xanthones (Bennett and Lee, 1989; Peters et al., 1997). They can be further hydroxylated at multiple positions, for example, the hydroxylation at C-6 is catalyzed by another P450, xanthone 6-hydroxylase (X6H) to yield 1,3,6,7- and 1,3,5,6- tetrahydroxyxanthones, respectively (Schmidt et al., 2000). In addition, xanthones can undergo stepwise prenylations and concomitant cyclizations to give complex derivatives with bridged polycyclic skeletons possessing high pharmaceutical potential (Zhang et al., 2004; Rukachaisirikul et al., 2005; Shadid et al., 2007).

On the other hand, in fungi and lichens, the xanthone nucleus originates exclusively from acetate. One molecule of acetyl-CoA is condensed with seven molecules of malonyl-CoA to form an octaketide intermediate that is subsequently cyclized to form an anthroquinone, followed by oxidative cleavage to give a carboxylated benzophenone intermediate (**Figure I-10**). The subsequent steps differ according to the organism; in some cases the resulting benzophenone intermediate is decarboxylated and cyclized directly into xanthones, while in other cases the cyclization yields a tetrahydroxanthone species followed by aromatization (Krohn et al., 2009; Masters and Bräse, 2012; Wezeman et al., 2015)

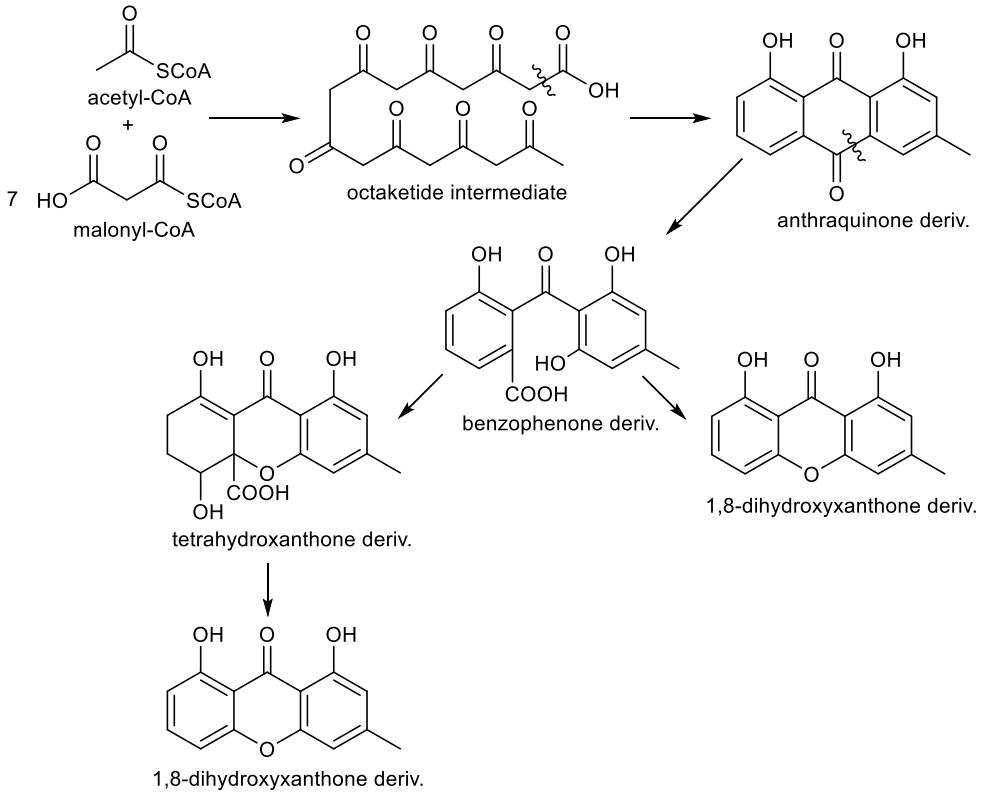


Figure I-10. Biosynthesis of xanthones in fungi and lichens.

3.3. Biological activities of xanthones

Xanthones exhibit diverse biological activities in a broad spectrum of diseases because of their interaction with a wide range of target biomolecules. Therefore, they are referred to as 'privileged structures' (Masters and Bräse, 2012; Winter et al., 2013). The spectrum of biological activities (**Figure I-11**) includes, but is not limited to, antimicrobial, antioxidant, anti-inflammatory, cholinesterase inhibitory and cytotoxic activities (Fotie and Bohle, 2006; El-Seedi et al., 2010).

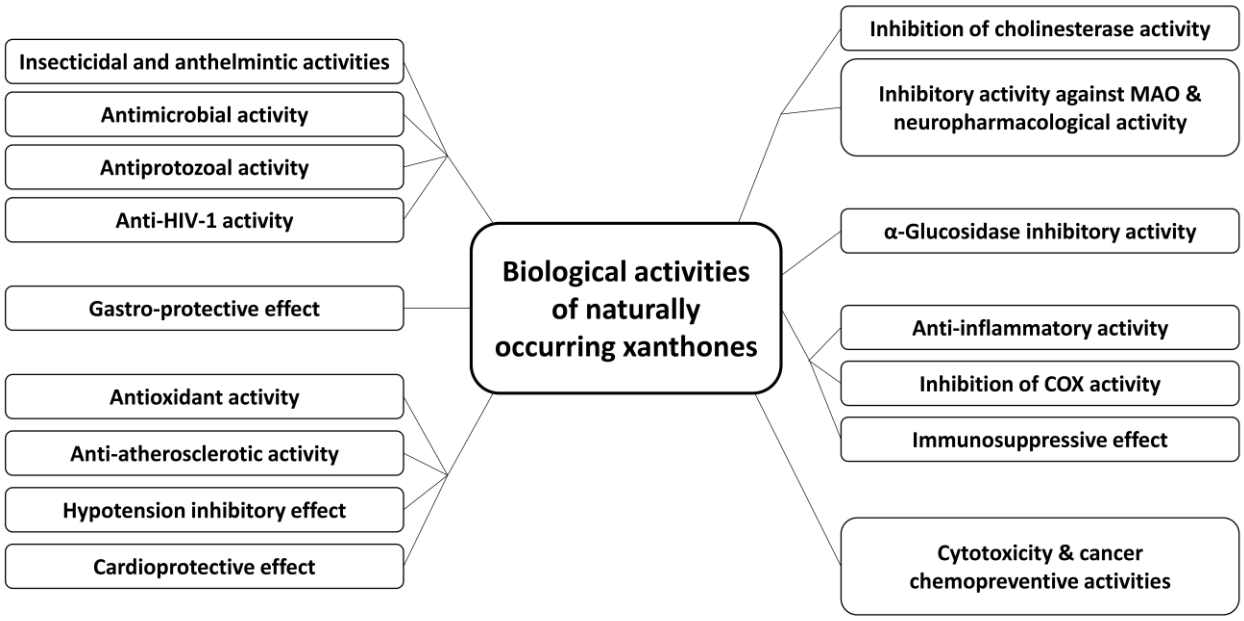


Figure I-11. Summary of the biological activities of naturally occurring xanthones; adapted from (El-Seedi et al., 2010).

3.3.1. Antimicrobial activity

3.3.1.1. Antibacterial activity

Several studies have shown that prenylated xanthone derivatives have pronounced antibacterial activities, especially towards resistant strains such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant enterococci (VRE) (Iinuma et al., 1996; Fukai et al., 2005; Rukachaisirikul et al., 2005). The bactericidal activity is suggested to be mediated through the interaction of the prenylated phenolic rings with the membrane lipid bilayer leading to disruption of the bacterial membrane and subsequently cell death (Koh et al., 2013).

3.3.1.2. Antiviral activity

Simple, prenylated as well as bis-xanthones including 1,7-dihydroxyxanthone (euxanthone), 1,3,8-trihydroxy-2,4-dimethoxyxanthone, mangostin, γ-mangostin and swertifrancheside showed anti-HIV-1 activity (Reutrakul et al., 2007). This activity could be due to either the HIV-protease or HIV-reverse transcriptase inhibitory properties of the mentioned xanthone derivatives (Wang et al., 1994; Chen et al., 1996).

3.3.1.3. Antiparasitic activity

Prenylated xanthones such as garciniaxanthone-A and B exhibit antiprotozoal activity against *Plasmodium*, *Trypanosoma* and *Leishmania* species (Abe et al., 2003; Hay et al., 2004; Mbwambo et al., 2006; Molinar-Toribio et al., 2006; Al-Massarani et al., 2013). In addition, insecticidal and anthelmintic activities have been observed for the fungus-produced dimeric xanthone xanthonol. Therefore, they may be considered promising leads for the development of new safer and more effective systemic antiparasitic drugs (Ondeyka et al., 2006).

3.3.2. Antioxidant activity

Xanthones, like many other phenolic compounds, show pronounced antioxidant activities. They act as free radical scavengers as shown by the results of the DPPH (1,1-diphenyl-2-piccrylhydrazyl) test performed on various naturally occurring xanthones (Deachathai et al., 2005; Mahabusarakam et al., 2006). This activity is beneficial in combating several diseases, in which oxidative stress and generation of ROS (reactive oxygen species) play a major role. These include neurodegenerative diseases such as Parkinson's and Alzheimer's diseases (Martinez et al., 2012; Wang et al., 2012).

3.3.3. Cardiovascular activities

Xanthones have beneficial effects on various cardiovascular ailments such as hypertension, thrombosis, atherosclerosis and ischemic heart disease (Jiang et al., 2004; El-Seedi et al., 2010). Cudraxanthone-D and isocudraniaxanthone-B are examples of xanthones showing ani-atherosclerotic activity (Park et al., 2006). This activity is attributed to their acyl CoA:cholesterol acyltransferase (ACAT) inhibitory and anti-low density lipoprotein (LDL) oxidation activities. Gentiacaulein and gentiakochianin induce vasodilation in rat aorta (Chericoni et al., 2003). A simple tetraoxygenated xanthone, namely 1-hydroxy-2,3,5-trimethoxyxanthone, caused a concentration-dependent potent relaxation in rat coronary artery (Wang et al., 2007).

3.3.4. Anti-inflammatory activity

γ-Mangostin has been demonstrated to have potent inhibitory activity on both inducible and constitutive cyclooxygenase (COX) enzymes (Nakatani et al., 2002). *In vivo* studies showed that α-mangostin had an inhibitory effect on carrageenan-induced paw edema in mice (Chen et al., 2008).

3.3.5. Cytotoxic activity

Xanthones are promising molecules as anti-cancer agents. The prenylated xanthones α-, β- and γ-mangostin showed inhibitory effect on cell growth in human colon cancer cells. In addition, the antiproliferative effect was accompanied by apoptosis (Matsumoto et al., 2003). Gambogic acid has a potent cytotoxic activity against a wide array of tumor cells (Felth et al., 2013; Yang and Chen, 2013).

3.3.6. Xanthones as promising agents for the treatment of Alzheimer's disease

Alzheimer's is a neurodegenerative disease characterized by gradual, irreversible memory loss leading to dementia, personality changes, unusual behavior and finally death. The etiology involves multiple factors that lead to the accumulation of amyloid-β peptide and tau protein in the brain neurons, increased lipid peroxidation and oxidative stress and decreased acetylcholine levels. Xanthones possess several properties which may be beneficial in combating the condition. Inhibition of acetylcholinesterase, free radical scavenging and MAO inhibitory activities have been observed for xanthones. Optimizing a balance between activities in one xanthone-based molecule would be of great benefit in the one-drug-multiple-target strategy for the treatment of Alzheimer's (Mattson, 2004; Zhang, 2005; Wang et al., 2012).

4. Aim of the work

Regioselective oxidative phenol coupling reactions have been established on the biochemical level to be responsible for the cyclization of 2,4,6-trihydroxybenzophenone into 1,3,7- and 1,3,5-trihydroxyxanthones in *Hypericum androsaemum* and *Centaurium erythraea*, respectively (Peters et al., 1997). The reactions are catalyzed by two distinct enzymes known as trihydroxyxanthone synthases (TXSs). They were demonstrated to be cytochrome P450 enzymes, as inferred from their NADPH and atmospheric oxygen dependencies. Furthermore, the cyclization reactions were hindered by common CYP inhibitors (Peters et al., 1997). However, no cDNA encoding either of those enzymes has been cloned so far and, hence, those xanthone synthases have never been studied at the molecular level.

The focus of interest of this work is to clone and functionally express the xanthone synthases from *Hypericum sp.* and to study their properties. A molecular biological approach has been chosen as isolation of CYPs was difficult due to their low expression levels, instability and membrane integration nature. The approach is based on an established *H. calycinum* cell culture system, which produces a derivative of 1,3,7-trihydroxyxanthone exclusively upon yeast elicitation (Gaid et al., 2012). A suppression subtractive hybridization (SSH) library was established from this system and is expected to contain high copy numbers of the transcripts encoding 1,3,7-trihydroxyxanthone synthase (1,3,7-TXS). The following steps should be performed in order to isolate and functionally characterize 1,3,7-TXS.

- Bioinformatic analysis of the expressed sequence tags (ESTs) of the subtracted library in order to identify candidate core fragments that encode CYPs.
- Completion of the coding sequence of the candidate fragments by rapid amplification of cDNA ends (RACE) technology to obtain full-length cDNA(s).
- As CYP activity requires the presence of an electron donor, a similar strategy would follow to clone a full-length cDNA encoding NADPH-dependent cytochrome P450 reductase (CPR).
- Expression of both enzymes in yeast cells and testing the activity of microsomal fractions with the potential substrates.

In order to isolate the gene encoding 1,3,5-trihydroxyxanthone synthase (1,3,5-TXS) the following strategy should be followed.

- Searching the public transcriptomic database of *H. perforatum* for candidate sequences sharing high sequence identity with the isolated 1,3,7-TXS cds.
- Expression of the isolated genes together with CPR in yeast cells and testing the microsomal fraction with the potential substrates.

Once both variants are successfully identified, a modelling-based reciprocal mutagenesis approach should be used to identify their regioselectivity determinants and convert each enzyme to its counterpart.

II. Materials

1. Biological materials

1.1. Plant materials

Hypericum perforatum plants were grown in the medicinal plants garden of the Institute of Pharmaceutical Biology, TU Braunschweig. Leaves were collected during the flowering stage, snap-frozen in liquid nitrogen and stored in -80 °C for subsequent use.

1.2. Cell cultures

Hypericum calycinum (Hypericaceae) cell cultures were grown in the dark at 25 °C in 50 ml liquid LS medium (Linsmaier and Skoog, 1965) with shaking at 120 rpm. Cultures were subdivided every 14 days in fresh LS media (3 g cells in 50 ml media).

1.3. Bacterial and yeast strains

Escherichia coli

Strain name	Genotype
DH5α	*F′ φ80δlacZΔM15 end A1 rec A1 hsdR17(rk⁻ mk⁺) supE44 thi-1 λ⁻ gyrA96 relA1 Δ(lacZYA-argFV169) deoR*

Saccharomyces cerevisiae

Strain name	Genotype
INV*Sc*1	*MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52*

2. Vectors

pGEM®-T Easy vector (Promega) was used for TA cloning of PCR products
pESC-URA vector (Agilent) were used for functional expression in yeast cells

3. Primers

3.1. Primers used for core fragments amplification, 3′ and 5′ RACE

Primer name	Sequence 5′→3′
Core fragments amplification:	
Contig59-F1	AAGGTCCCCGTCCTCGTTGTC
Contig21-R1	CGTATTCCTCGAACCTCTC
CPR-dpF-1	GCNGARGGNTTYGCHAAGGC
CPR-R1	CTTGGACAATGGTGTGGAGAGTTC
3′ and 5′ RACE:	
CYP81-3RACE1	TCCTAATGCAACACCCAGAG
CYP81-3RACE2	CTACCTAAGGGCCGTCGTGA
CYP81-5RACE1	CCGCTTCTCCTCCTTGGTGTCC
CYP81-5RACE2	CGTCGTTCTTGGTGAAGCACTCCTC
CPR-3RACE1	GCTCCATTCAGGGGTTTCC
CPR-3RACE2	GGTTAGCCCTGAAAGAATCC
CPR-5RACE1	ATCCATTTCGGGCCACAATGCCTCTC
CPR-5RACE2	TGCCTGTTCCCTAAGCCAAACACTGC
51544-3RACE1	GACGCCATGCGACTTCATCC
51544-3RACE2	TCCTTAACGGCTTACTTGCCAATA

3.2. Primers used for cloning

Primer name	Sequence 5′→3′ (Introduced restriction site sequences are underlined)
HcCPR2-*Bam*HI-F	GTAATGGATCCGATGGAACCGACGGGGAGC
HcCPR2-*Hin*dIII-R	GTCTGAAGCTTTCACCATACGTCGCGAAGGTAC
Hc81AA1-*Eco*RI-F	ATTGAATTCGGATGGAGGACTTGTACTTGTACC
Hc81AA1-*Pac*I-R	ACGTTAATTAATTAGAGCTGGGAAAGGAGACC
Hp81AA1-*Spe*I-F	TATGACTAGTAATGGAGGACTTGTATCTGTACG
Hp81AA1-*Pac*I-R	ACGTTAATTAATTAGAGCTGGGAGAGGAGG
Hp81AA2-*Eco*RI-F	ATTGAATTCATGGACGGTTTATACTTAAAACTAGCC
Hp81AA2-*Pac*I-R	ACGTTAATTAACTAGAGTTGGGAAAGGAGCTTG
Hp81AA2-*Spe*I-F	ATTACTAGTAATGGACGGTTTATACTTAAAACTAGC
Hp81AA3-*Spe*I-F	TATGACTAGTAATGGAATTGTATTTGTATCTAGCCG
Hp81AA3-*Pac*I-R	ACGTTAATTAACTAAGAGAGTAGAGAACTGAGG

3.3. Primers used for C-terminal exchanges and site-directed mutagenesis

Primer name	Sequence 5′→3′
Overlapping primers used for C-terminal exchanges:	
AA1-AA2-ex	CGGAGGGATGGGGACCCACATTGTCTGCGAGACATTGGGAATG
AA2-AA1-ex	CCGGAATGGGAACCCATATCTCGTCCCTGGGGATTGGAATTCTTA
Primers used for site-directed mutagenesis (mutated positions are highlighted in red):	
AA1-A373S-F	CCGCCGTCCCCACTCTCCCTCCCCCACTTCTCCAACCAAG
AA1-A373S-R	GTGGGGACGGCGGGTACATCCTGAGCGTCTCCTTCACGAC
AA1-ASSL-F	CGCCGTCCCCACTCTTGCTCCCCCACTTCTCCAACCAAG
AA1-T482V-F	AGCCACGCGCGTGGTGTCGCGAGTCTGTCAAGGAAGAAG
AA1-T482F-F	AGCCACGCGCGTGGTTTCGCGAGTCTGTCAAGGAAGAAG
AA1-T482uni-R	ACGCGCGTGGCTCATGTCCTCGTCCATGCCGAGGTGGC
AA1-T482A-F	CGTGGTGCCGCGAGTCTGTCAAGGAAGAAGCCATTGGAG
AA1-T482A-R	TCGCGGCACCACGCGCGTGGCTCATGTCCTCGTCCATGC
AA1-TALMRKKN-F	CGTGGTGCCGCGAGTATGTCAAAGAACAAGCCATTGGAG
AA1-VIMT-F	GAATTATTACGATGATGTTTATTGCCGGGACCGAGACG
AA1-VIMT-R	TCATCGTAATAATTCCCTTGACGATGTCATCCGTG
AA1-W113F-F	TCGTGTTCGCCCCGCACGGCCCCATCTGGAAGAGCCTC
AA1-W113F-R	GGGGCGAACACGAGGAAGGTGTAGTCGTAGGTGAGGTGC
AA1-I220P-F	ATGACCCCATGCGACTTCATCCCGGTGATGAGGCTGATC
AA1-I220P-R	TCGCATGGGGTCATGGAGACGATAGGAAAGAAGAGATCC
AA2-A483T-F	GCCGGTACGGCTAGTATGTCCAAGAACAAGCCATTGG
AA2-A483T-R	TAGCCGTACCGGCTTGGTGGGTCATGTCCTCGTCCATAC
AA2-ATMLKRNK-F	GCCGGTACGGCTAGTCTGTCCAGGAAGAAGCCATTGG
AA2-L378S-F	CGCTCTCGCTACCTCGTTATTCGAGCGAGGCTTGCAC
AA2-L378S-R	GGTAGCGAGAGCGGGGACGGAGGATACAGCCTCATCGTC
AA2-S375A-F	CCTCCGGCCCCGCTCTTGCTACCTCGTTATTCGAGCGAG
AA2-S375A-R	GCGGGGCCGGAGGATACAGCCTCATCGTCTCCTTCACG
AA2-SALS-F	CCTCCGGCCCCGCTCTCGCTACCTCGTTATTCGAGCGAG

3.4. SMART II RACE primers

Primer name	Sequence 5′→3′
SMART II™ oligo	AAGCAGTGGTATCAACGCAGAGTACGCGGG
3′-RACE CDS primer	AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$VN
5′-RACE CDS primer	$(T)_{25}$VN
RACE long	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACG CAGAGT
RACE short	CTAATACGACTCACTATAGGGC
RACE nested	AAGCAGTGGTATCAACGCAGAGT

3.5. Primers used for real-time qRT-PCR

Primer name	Sequence 5′→3′
Hc81AA1-qF	TCTCGATGACCGTATGTGACTT
Hc81AA1-qR	CGTTAAGGAAGACCTCCCTCT
HcCPR2-qF	TGACTATGCTGCGGATGATGAA
HcCPR2-qR	GCCAAGAAGAGGACAACCAAATC
HcBPS-qF	AAGGAAAGAAGAGGGCTAGTGT
HcBPS-qR	ATGTGCTCGCTGTTAGTGTTC
Actin-qF	CGGCAGTGGTTGTGAACAT
Actin-qR	TCTCGCTGGTCGTGATCTG
Histone-H2A-qF	AACATCTACTCTTTGGACGACTTG
Histone-H2A-qR	AATTGCTGGAGGTGGAGTTATTC

4. Enzymes

4.1. Enzymes used for reverse transcription (RT)

Enzyme name	Manufacturer
RevertAid™ HMinus Reverse Transcriptase	Thermo Scientific
SMARTScribe™ Reverse Transcriptase	Takara

4.2. Enzymes used for polymerase chain reaction (PCR)

Enzyme name	Manufacturer
peqGOLD *Taq* DNA polymerase	Peqlab
Phusion® Hot Start II High-Fidelity DNA Polymerase	Thermo Scientific
Q5® Hot Start High-Fidelity DNA Polymerase	New England Biolabs

4.3. Restriction endonucleases

Enzyme name	Sequence	Manufacturer
*Bam*HI	G^GATCC	Thermo Scientific
*Dpn*I	GA_m^6^TC	Thermo Scientific
*Eco*RI	G^AATTC	Thermo Scientific
*Hin*dIII	A^AGCTT	Thermo Scientific
*Pac*I	TTAAT^TAA	Thermo Scientific
*Spe*I	A^CTAGT	Thermo Scientific

4.4. Miscellaneous enzymes

Enzyme name	Manufacturer
FastAP Thermosensitive Alkaline Phosphatase	Thermo Scientific
RiboLock™ RNase Inhibitor	Thermo Scientific
RNase A	Thermo Scientific
RNase-free DNase I	Qiagen
T4 DNA ligase	Thermo Scientific
Instant Sticky-end Ligase Master Mix	NEB

5. Kits

Name	Purpose	Manufacturer
GeneJET™ Plant RNA Purification Mini Kit	RNA isolation from *H. perforatum* leaves	Thermo Scientific
RNeasy® Plant Mini Kit	RNA isolation from *H. calycinum* cell cultures	Qiagen
S.c. EasyComp™ Transformation Kit	Preparation and transformation of yeast competent cells	Invitrogen
innuPREP DOUBLEpure Kit	DNA purification from PCR, digestion mixture and agarose gel	Analytik Jena
iScript™ Reverse Transcription Supermix for RT-qPCR	Synthesis of first strand cDNA for real-time qRT-PCR	Bio-Rad
iTaq™ Universal SYBR® Green Supermix	Real-time qRT-PCR	Bio-Rad

6. Chemicals

Chemical	Manufacturer
1,3-Dihydroxyxanthone	Lab collection
1-Naphthylacetic acid (NAA)	Fluka
2,2′,3,4′-Tetrahydroxybenzophenone	Prof. You's lab, China
2,2′,4,5′-Tetrahydroxybenzophenone	Prof. You's lab, China
2,3,4-Trihydroxybenzoic acid	Acros
2,3′,4,4′,6-Pentahydroxynenzophenone (maclurin)	Aldrich
2,3′,4,6-Tetrahydroxybenzophenone	Lab collection
2,3-Dihydroxybenzoic acid	Acros
2,4,6-Trihydroxybenzophenone	ICN Biomedicals
2,4-Dichlorophenoxyacetic acid (2,4-D)	Fluka
2,4-Dihydroxybenzophenone	Fluka
2,5-Dihydroxybenzoic acid	Acros
2-Hydroxybenzophenone	Merck
2-Mercaptoethanol (β-ME)	Roth
4-Hydroxybenzophenone	Fluka
Acetic acid	Roth
Acrylamide/Bisacrylamide 30%	Bio-Rad
Agar Kobe	Applichem

Continued:

Chemical	Manufacturer
Ammonium persulfate (APS)	Roth
Ampicillin	Roth
Bacto casamino acids	Difco
Betaine	Sigma-Aldrich
Bovine serum albumin (BSA)	Sigma-Aldrich
Bromophenol blue	Aldrich
$CaCl_2.2H_2O$	Roth
Chloroform	Sigma-Aldrich
$CoCl_2.6H_2O$	Riedel-de Haën
Coomassie blue (G250 and R250)	Merck
$CuSO_4.5H_2O$	Riedel-de Haën
Ethanol	VWR
Ethidium bromide solution (1%)	Roth
Ethyl acetate	VWR
$Na_2EDTA.2H_2O$	Roth
$FeSO_4.7H_2O$	Riedel-de Haën
Formic acid	Roth
Galactose	Roth
Glass beads	Sigma-Aldrich
Glucose	Roth
Glycerol	Roth
Glycine	Roth
H_3BO_3	Roth
HCl	VWR
IPTG	Applichem
Isopropanol	VWR
KCl	Roth
KH_2PO_4	Roth
KI	Riedel-de Haën
KNO_3	Roth
Methanesulfonic acid	Sigma-Aldrich
$MgCl_2.6H_2O$	Fluka
$MgSO_4.7H_2O$	Roth
Midori Green	Nippon Genetics GmbH
$MnSO_4.H_2O$	Riedel-de Haën
Myo-inositol	Roth
$Na_2MoO_4.2H_2O$	Merck
NaCl	Roth
NaOH	Roth
NH_4NO_3	Roth
o-Phosphoric acid	Fischer-Scientific
PEG 4000	Fluka
Peptone from casein	Roth
peqGold Universal Agarose	Peqlab

Continued:

Chemical	Manufacturer
Phenol	Roth
Phloroglucinol	Acros
Phosphorus pentoxide	Acros
Polyvinylpyrrolidone	Sigma-Aldrich
Potassium acetate	Roth
Sodium dodecyl sulfate (SDS)	Roth
Sorbitol	Roth
Sucrose	Diadem
Tetramethylethylenediamine (TEMED)	Roth
Thiamine hydrochloride	Sigma-Aldrich
tris(hydroxymethyl)aminomethane (Tris)	Roth
X-Gal	Sigma-Aldrich
Xylene cyanol FF	Roth
Yeast extract	Applichem
Yeast nitrogen base without amino acids	Sigma
ZnSO4.7H2O	Riedel-de Haën

7. Culture media

7.1. Plant cell culture medium

LS medium				
Macroelements	**10X stock solution**			
NH_4NO_3	16.5	g/l	1650.0	mg/l
KNO_3	19.0	g/l	1900.0	mg/l
$MgSO_4.7H_2O$	3.7	g/l	370.0	mg/l
KH_2PO_4	1.7	g/l	170.0	mg/l
$CaCl_2.2H_2O$	4.4	g/l	440.0	mg/l
$Na_2EDTA.2H_2O$	413.0	mg/l	41.3	mg/l
$FeSO_4.7H_2O$	278.0	mg/l	27.8	mg/l
Microelements	**1000X stock solution**			
$MnSO_4.H_2O$	16.9	g/l	16.9	mg/l
$ZnSO_4.7H_2O$	10.6	g/l	10.6	mg/l
KI	830.0	mg/l	830.0	µg/l
H_3BO_3	6.2	g/l	6.2	mg/l
$Na_2MoO_4.2H_2O$	250.0	mg/l	250.0	µg/l
$CuSO_4.5H_2O$	25.0	mg/l	25.0	µg/l
$CoCl_2.6H_2O$	25.0	mg/l	25.0	µg/l
Vitamins	**100X stock sloution**			
Thiamine hydrochloride	4.0	mg/100ml	0.4	mg/l
Myo-inositol	1.0	g/100ml	100.0	mg/l
Hormones	**Stock solution**			
2,4-Dichlorophenoxyacetic acid (2,4-D)	1.0	mg/ml	220.0	µg/l
1-Naphthylacetic acid (NAA)	1.0	mg/ml	186.0	µg/l
Sucrose			30.0	g/l

- The pH is adjusted to 6.0 and the medium is autoclaved.

7.2. Bacterial culture media

LB medium			Final concentration	
Yeast extract	5.0	g/l	0.5	% (w/v)
Peptone from casein	10.0	g/l	1.0	% (w/v)
NaCl	10.0	g/l	1.0	% (w/v)

- The pH is adjusted to 7.0 with 5N NaOH and the medium is autoclaved.
- For solid medium, 2% (w/v) agar is added before autoclaving.
- For selection via ampicillin resistance, sterile filtered ampicillin is added after autoclaving at a final concentration of 100 mg/l.
- For blue-white screening, ampicillin (100 mg/l), IPTG (12 mg/l) and X-GAL (40 mg/l, dissolved in DMF) are added.

SOC medium			Final concentration	
Yeast extract	5.0	g/l	0.5	% (w/v)
Peptone from casein	20.0	g/l	2.0	% (w/v)
NaCl	467.6	mg/l	10.0	mM
KCl	186.4	mg/l	2.5	mM
$MgSO_4$. $7H_2O$ (1M)	5.0	ml/l	5.0	mM
$MgCl_2$ (1M)	5.0	ml/l	5.0	mM
Dextrose (1M)	20.0	ml/l	20.0	mM

- The pH is adjusted to 7.5.
- The first four components are dissolved in 970 ml distilled water and autoclaved. Afterwards, the sterile-filtered solutions of the last three components are added.

7.3. Yeast culture media

YPD medium			Final concentration	
Yeast extract	10.0	g/l	1.0	% (w/v)
Peptone from casein	20.0	g/l	2.0	% (w/v)
Dextrose	20.0	g/l	2.0	% (w/v)

- The first two components are dissolved in 900 ml distilled water and autoclaved. Afterwards, 100 ml of sterile-filtered 20% (w/v) dextrose solution are added.
- For solid medium, 2% (w/v) agar is added before autoclaving.

Synthetic dextrose minimal medium (SD dropout medium)			Final concentration	
Yeast nitrogen base (with or without amino acids)	6.7	g/l	0.67	% (w/v)
Bacto casamino acids	1.0	g/l	0.10	% (w/v)
Dextrose	20.0	g/l	2.0	% (w/v)

- The first two components are dissolved in 900 ml distilled water and autoclaved. Afterwards, 100 ml of sterile-filtered 20% (w/v) dextrose solution are added.
- For solid medium, 2% (w/v) agar is added before autoclaving.

YPGE medium			Final concentration	
Yeast extract	10.0	g/l	1.0	% (w/v)
Peptone from casein	10.0	g/l	1.0	% (w/v)
Dextrose	5.0	g/l	0.5	% (w/v)
Ethanol	30.0	ml/l	3.0	% (v/v)

- The first two components are dissolved in 945 ml distilled water and autoclaved. Afterwards, 25 ml of sterile-filtered 20% (w/v) dextrose solution and 30 ml absolute ethanol are added.

8. Buffers and solutions

Unless otherwise indicated, all buffers and solutions are autoclaved before first use.

8.1. Buffers used for plasmid isolation from *E. coli*

Buffer I			**Final concentration**	
Tris	1.5	g/250 ml	50	mM
$Na_2EDTA.2H_2O$	931	mg/250 ml	10	mM
RNase A (added immediately before use)			10	µl/ml

- The pH is adjusted to 8 with 1N HCl

Buffer II			**Final concentration**	
NaOH	2	g/250 ml	0.2	M
SDS	2.5	g/250 ml	1	% (w/v)

Buffer III			**Final concentration**	
Potassium acetate	73.6	g/250 ml	3	M

- The pH is adjusted to 5.5 with glacial acetic acid (requires about 55 ml)

8.2. Buffer used for plasmid isolation from yeast (Hoffman and Winston, 1987)

Yeast cracking solution			**Final concentration**	
Tris	60.6	mg/50 ml	10	mM
$Na_2EDTA.2H_2O$	18.6	mg/50 ml	1	mM
NaCl	292	mg/50 ml	100	mM
SDS	500	mg/50 ml	1	% (w/v)
Triton X-100	1	ml/50 ml	2	% (v/v)

- The pH is adjusted to 8.5 with 1N HCl

8.3. Buffers used for microsomal protein isolation from yeast

1M Tris-HCl pH 7.4 stock buffer			**Final concentration**	
Tris	12.1	g/100 ml	1	M

- The pH is adjusted to 7.4 with 1N HCl

TEK buffer			**Final concentration**	
Tris-HCl pH 7.4 (1M stock buffer)	12.5	ml/250 ml	50	mM
$Na_2EDTA.2H_2O$	93	mg/250 ml	1	mM
KCl	1.864	g/250 ml	0.1	M

TES buffer			**Final concentration**	
Tris-HCl pH 7.4 (1M stock buffer)	25	ml/500 ml	50	mM
$Na_2EDTA.2H_2O$	186.1	mg/500 ml	1	mM
Sorbitol	54.651	g/500 ml	0.6	M

TEG buffer			Final concentration	
Tris-HCl pH 7.4 (1M stock buffer)	5	ml/100 ml	50	mM
$Na_2EDTA.2H_2O$	73	mg/100 ml	1	mM
Glycerol	20	ml/100 ml	20	% (v/v)

8.4. Buffers and solutions used for agarose gel electrophoresis

50X TAE buffer			Final concentration	
Tris	242.28	g/l	2	M
$Na_2EDTA.2H_2O$	18.612	g/l	50	mM

- The pH is adjusted to 8.0 with glacial acetic acid

6X DNA loading dye			Final concentration	
Tris	60.6	mg/50 ml	10	mM
$Na_2EDTA.2H_2O$	1.117	g/50 ml	60	mM
Bromophenol blue	15	mg/50 ml	0.03	% (w/v)
Xylene cyanol FF	15	mg/50 ml	0.03	% (w/v)
Glycerol	30	ml/50 ml	60	% (v/v)

- The pH is adjusted to 7.6 with 1N HCl.
- Stored in -20 °C without autoclaving.

8.5. Buffers and solutions used for SDS-PAGE

5% Staking gel (for 2 gels)		
Water	2.72	ml
Tris-HCl 1M pH 6.8	504	µl
Acrylamide/Bisacrylamide 30% (w/v)	664	µl
SDS 10% (w/v)	40	µl
APS 10% (w/v)	40	µl
TEMED	4	µl

12% Resolving gel (for 2 gels)		
Water	2.3	ml
Tris-HCl 1.5M pH 8.8	1.75	ml
Acrylamide/Bisacrylamide 30% (w/v)	2.8	ml
SDS 10% (w/v)	70	µl
APS 10% (w/v)	70	µl
TEMED	2.8	µl

2X Protein loading buffer		
Distilled water	2.7	ml
Tris-HCl 0.5M pH 6.8	1.0	ml
SDS 10% (w/v)	3.3	ml
Bromophenol blue 0.5% (w/v)	0.5	ml
Glycerol	2.0	ml
β-Mercaptoethanol (Added immediately before use)	0.5	ml

10X Electrophoresis buffer			Final concentration	
Tris	15	g/500 ml	80.6	mM
Glycine	72	g/500 ml	1.9	M
SDS	5	g/500 ml	1	% (w/v)

Coomassie blue R250 stock solution			Final concentration	
Coomassie blue R250	500	mg	1	% (w/v)
Methanol	50	ml		

- The solution is filtered before further use.

Staining solution			Final concentration	
Coomassie blue (stock solution)	20	ml/200 ml	0.1	% (w/v)
Acetic acid	20	ml/200 ml	10	% (v/v)
Methanol	100	ml/200 ml	50	% (v/v)

Destaining solution			Final concentration	
Acetic acid	20	ml/200 ml	10	% (v/v)
Methanol	30	ml/200 ml	15	% (v/v)

8.6. Solution for the determination of protein concentration

Bradford dye solution			Final concentration	
Coomassie blue G250	100	mg/l	0.01	% (w/v)
Ethanol absolute	50	ml/l	5	% (v/v)
o-Phosphoric acid (85%)	100	ml/l	10	% (v/v)

- Coomassie blue is dissolved first in ethanol and then *o*-phosphoric acid is added.
- The solution is filtered before use and kept at 4 °C away from light.

8.7. Combinatorial enhancer solution for PCR (Ralser et al., 2006)

5X CES			Final concentration	
Betaine	3.2	g/10 ml	2.7	M
BSA (1 mg/ml stock solution)	550	µl/10 ml	55	µg/ml
DMSO	670	µl/10 ml	6.7	% (v/v)
DTT	10.3	mg/10 ml	6.7	mM

9. Equipment

9.1. General equipment

Equipment	Model	Manufacturer
Autoclave	Systec VX-120	Systec
Balance	2254	Sartorius
Balance	Kern 572	Kern & Sohn Gmbh
Balance	LA230S	Sartorius
Centrifuge	Biofuge pico	Heraeus
Centrifuge	Avanti J-E	Beckmann Coulter
Centrifuge	Avanti J-30I	Beckmann Coulter

Continued:

Equipment	Model	Manufacturer
Centrifuge	Universal 32 R	Hettich
Centrifuge	Biofuge 13	Heraeus
Deep freezer (-80 °C)	MDF-U53V	Sanyo
Dry block heater	Dri-Block DB-3D	Techne
Electrophoresis chamber for agarose gels	Wide Mini-Sub Cell GT	Bio-Rad
Electrophoresis chamber for polyacrylamide gels	Mini PROTEAN Tetra Cell	Bio-Rad
Gel documentation	Infinity-3000	Vilber Lourmat
Heating circulator water bath	TopTech MW-4	Julabo
Laminar air flow	LaminAir HLB2472 BS	Heraeus
Laminar air flow	LaminAir HBB2460	Heraeus
Microscope	Laborlux 11	Leitz
pH meter	pH 325	WTW GmbH
Real-Time PCR Detection System	CFX Connect™	Bio-Rad
Rotary evaporator	W60	Heidolph
Rotational Vacuum Concentrator	RVC 2-18	Christ
Shaker incubator	Multitron	Infors HT
Shaker incubator	KF4	Infors HT
Spectrophotometer	Ultrospec 1000	Pharmacia Biotech
Spectrophotometer	Ultrospec 3100 pro	GE Healthcare
Spectrophotometer	Cary 300 UV-Vis	Varian
Spectrophotometer	SimpliNano™	GE Lifesciences
Thermocycler	T-Professional basic gradient	Biometra
Diaphragm vacuum pumps with programmable digital vacuum controller	VCZ 224	Ilmvac
Vortex shaker	VF2	IKA
Water purification system	Arium 611	Sartorius

9.2. Equipment for HPLC and MS analysis

Equipment	Model	Manufacturer
HPLC	**Agilent 1260 Infinity System**	Agilent
Pump	G1311C Quaternary Pump VL	Agilent
Auto sampler	G1329B ALS	Agilent
Column oven	G1316A TCC	Agilent
Column	HyperClone ODS 120A (C18, 4.6 × 150 mm, 5 µ)	Phenomenex
Diode Array Detector	G1315D DAD VL	Agilent
Software	Agilent ChemStation for LC 3D System Rev. B.04.03(16)	Agilent

Continued:

Equipment	Model	Manufacturer
HPLC	**LaChrom Elite System**	VWR-Hitachi
Pump	L-2130	VWR-Hitachi
Degasser	Model 2005 Degasser	VWR-Hitachi
Auto sampler	L-2200	VWR-Hitachi
Column	ZORBAX Eclipse Plus (C18, 4.6 × 100 mm, 3.5 µ)	Agilent
Diode Array Detector	L-2455	VWR-Hitachi
Software	EZChrom Elite vers. 3.3.2 SP1	Agilent
Mass spectrometer	**3200 QTRAP LC/MS/MS System**	AB Sciex
Software	Analyst vers. 1.4.2	AB Sciex

10. Software, online databases and tools

Name, web address*	Use
Software:	
Chromas Lite 2.1.1 http://technelysium.com.au/	Visualization of the sequencing chromatograms from .ab1 files.
Lasergene - DNAStar 7.0	A package composed of various individual programs for DNA and protein sequence analysis. Those programs include: MegAlign: generates pairwise and multiple sequence alignments of protein or DNA showing consensus strength and percent identity matrices. EditSeq: helpful in editing protein and nucleotide sequences. SeqMan: assembles Contigs from individual overlapping sequences. SeqBuilder: generates plasmid maps, shows existing restriction sites within a given nucleotide sequence and predicts restriction digestion results.
Beacon Designer 8.0	Design of primers for qRT-PCR experiments.
Primer premier 5.0	Design of primers for RACE experiments.
Bio-Rad CFX Manager 3.1	Design of qPCR plates, calculation of the amplification efficiencies and analysis of the data.
MEGA 6.0 http://www.megasoftware.net/	Multiple sequence alignment and construction of phylogenetic trees (Tamura et al., 2013).
YASARA View 15.9.6 http://www.yasara.org/index.html	Visualization of the three dimensional structure of macromolecules.

Continued:

Name, web address	Use, description
Online databases:	
NCBI protein and nucleotide databases http://www.ncbi.nlm.nih.gov/protein http://www.ncbi.nlm.nih.gov/nuccore	Retrieving amino acid and nucleotide sequences of deposited CYPs and CPRs.
MPGR database http://medicinalplantgenomics.msu.edu/	A repository of the transcriptomic and metabolomic data of 14 medicinal plants, including *H. perforatum*.
OneKP database http://www.onekp.com/	Transcriptomic data of over 1000 plants including *H. perforatum*, *Garcinia oblongifolia*, *G. livingstonei* and *Mammea americana*.
UIC transcriptome database http://apps.pharmacy.uic.edu/depts/pcrps/MedTranscriptomePlants/index.html	Transcriptome Characterization of Medicinal Plants Relevant to Human Health including *Calophyllum antillanum*.
PhytoMetaSyn database http://www.phytometasyn.ca/index.php	Transcriptomic data of 75 plant species including *H. perforatum*.
Online tools:	
BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi	blastn: Search a nucleotide database using a nucleotide query. blastp: Search protein database using a protein query. blastx: Search protein database using a translated nucleotide query in all possible frames. tblastn: Search translated nucleotide database using a protein query.
ExPASy translate tool http://web.expasy.org/translate/	Translation of DNA sequences in all 6 possible forward and reverse frames.
CAP3 http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::cap3	Assembling the ESTs from the subtractive library into Contigs.
Blast2GO https://www.blast2go.com/	Functional annotation of the assembled Contigs.
Primer 3.0 http://primer3.ut.ee/	Rechecking the properties of the designed qPCR primers.
ESPript 3.0 http://espript.ibcp.fr/ESPript/ESPript/	Display of sequence similarities and secondary structure elements of aligned sequences for analysis and publication purpose.
CYtochrome P450 Engineering Database (CYPED 6.0) https://cyped.biocatnet.de/	Blastable database of all named CYP sequences, tool for CYP standard numbering and prediction of secondary structure elements.

* All listed websites have been successfully accessed on October 21, 2015

III. Methods

1. Elicitor treatment of *Hypericum calycinum* cell cultures

Hypericum calycinum cell suspension cultures (II.1.2) were treated with of yeast extract as an elicitor to induce the production of a 1,3,6,7-tetrahydroxyxanthone derivative known as hyperxanthone E (Gaid et al., 2012), whereby all transcripts encoding the enzymes involved in its biosynthesis should be overexpressed.

At the 4th day after subdivision of the cultures, 1 ml of sterile-filtered yeast extract (150 mg/ml) were added to 50 ml cultures to a final concentration of 3 g/l. The cells were further incubated for 7 h and the cells were harvested by vacuum filtration and frozen in liquid nitrogen for subsequent RNA extraction.

2. Molecular biology methods

2.1. Isolation of nucleic acids

2.1.1. Isolation of total RNA

2.1.1.1. Isolation of total RNA from *Hypericum calycinum* cells

H. calycinum cell culture were elicited with 3 g/l yeast extract, cells were collected 7 h post elicitation. Total RNA was extracted with RNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol.

2.1.1.2. Isolation of total RNA from *Hypericum perforatum* leaves

H. perforatum middle-aged leaves were collected and total RNA was extracted using GeneJET Plant RNA Purification Mini Kit (Thermo Scientific) following the manufacturer's protocol for polyphenol-rich samples.

2.1.2. Isolation of plasmid DNA from *E. coli*

Plasmids were isolated from *E. coli* by the alkaline lysis method described by Birnboim and Doly (1979). This method depends on the selective denaturation of the large linear chromosomal DNA in alkaline medium with retaining the small circular plasmid DNA unharmed. After neutralization, chromosomal DNA and proteins precipitate, while plasmid DNA remains in the supernatant.

A single *E. coli* colony was inoculated in a 5 ml LB tube supplemented with 10 µl of 100 mg/ml sterile filtered ampicillin solution to positively select for bacterial cells harboring a plasmid with an ampicillin carrying resistance gene. The culture was incubated overnight (16–18 h) at 37 °C and 200 rpm. The cells in 2 ml of the overnight culture were pelleted in a 2 ml Eppendorf tubes by centrifugation for 5 min at 5000 rpm. The supernatant was discarded and the pellet was resuspended in 300 µl of ice-cooled buffer I to which RNase A was freshly added (II.8.1). The alkaline lysis was achieved by the addition of 300 µl of buffer II (II.8.1), followed by gentle mixing by flipping the tubes 4–6 times and incubation at room temperature for 5 min. Neutralization by the addition of 300 µl of buffer III (II.8.1), followed by gentle mixing and incubation on ice for 20 min resulted in the precipitation of chromosomal DNA and proteins. The white precipitate was discarded after centrifugation for 10 min at 13000 rpm, while 800 µl of the supernatant were vortexed with an equal volume of chloroform; the two phases were separated by centrifugation for

10 min at 13000 rpm. 700 µl of the aqueous phase were carefully transferred to a 1.5 ml Eppendorf and the plasmid DNA was precipitated by the addition of 490 µl isopropanol (70 % of the aqueous phase volume), followed by vortexing and centrifugation for 35 min at 13000 rpm. The supernatant was discarded and the pellet was washed with 500 µl of 70 % ethanol, centrifuged for 10 min at 13000 rpm and the supernatant was again discarded. The pelleted plasmid was dried in a rotational vacuum concentrator and finally dissolved in 30 µl of deionized water.
The isolated plasmids were checked by restriction digestion and/or by amplification of the inserts before sequencing. They could be used for subsequent *E. coli* or yeast transformations.

2.1.3. Isolation of plasmid DNA from *Saccharomyces cerevisiae*

Plasmid isolation from yeast cells was performed according to the protocol described by Hoffman and Winston (1987). The purpose was to control if the picked colonies after transformation actually contained the introduced plasmids in the course of protein expression.
The cells in 2 ml culture were pelleted by centrifugation for 5 min at 5000 rpm, the supernatant was discarded and the cells were resuspended in the residual medium left. After addition of 200 µl of the yeast cracking solution (II.8.2), 200 µl of phenol:chloroform (1:1) and 300 mg of acid washed glass beads, the cells were vigorously vortexed for 2 min. The homogenate was centrifuged for 10 min at 13000 rpm and 250 µl of the aqueous phase were transferred to a 1.5 ml Eppendorf. The DNA was precipitated by the addition of 175 µl isopropanol (70 % of the aqueous phase volume), vigorously mixed and centrifugation for 35 min at 13000 rpm. The supernatant was discarded and the pellet was washed with 500 µl of 70 % ethanol, centrifuged for 10 min at 13000 rpm and the supernatant was again discarded. The pelleted plasmid was dried in a rotational vacuum concentrator and finally dissolved in 30 µl of deionized water.
The plasmids isolated by this method were suitable for transformation in *E. coli* and for checking of the inserts by PCR amplification. However, they were not pure enough to be checked by restriction digestion.

2.1.4. Isolation of DNA from agarose gels

The desired DNA bands were extracted from agarose gels (III.2.4) by solubilizing the gel and selectively adsorbing the DNA on a special silica membrane, washing with an ethanol-containing buffer and finally elution of the bound DNA with an aqueous buffer. This was performed using the innuPREP DOUBLEpure Kit (Analytik Jena) following the manufacturer's protocol. In brief, the cut band was solubilized in 600 µl of the gel extraction buffer, heated for 10 min at 50 °C. The solution was mixed with 50 µl of the binding optimizer solution and transferred to the separation column, centrifuged at for 1 min at 8000 rpm. The column was washed twice, each with 700 µl of LS buffer (containing ethanol), centrifuged for 1 min at 12000 rpm and finally dried by centrifugation for 2 min at 13000 rpm. The DNA was eluted by the addition of 10–35 µl of the elution buffer (or dH_2O instead), incubated for 1 min at room temperature and finally centrifuged for 2 min at 10000 rpm.

2.1.5. Isolation of DNA from digestion or dephosphorylation reaction mixtures

The DNA was isolated from digestion or dephosphorylation reaction mixtures using the same principle described in (III.2.1.4). The same innuPREP DOUBLEpure Kit (Analytik Jena) was also used. The reaction mixture was mixed with 500 µl of the binding buffer, transferred to the separation column, centrifuged at for 1 min at 12000 rpm. The column was washed once with 700 µl of LS buffer, centrifuged for 1 min at 12000 rpm. The DNA was eluted by the addition of 10–20 µl of the elution buffer (or dH_2O instead), incubated for 1 min at room temperature and finally centrifuged for 2 min at 10000 rpm.

2.2. Quantification of nucleic acids

The concentration of nucleic acids solutions could be determined spectrophotometrically by measuring the UV absorbance at 260 nm. 5 µl of the DNA or RNA sample solution are diluted to 500 µl with deionized water and the absorbance at 260 and 280 nm were measured and the concentration was calculated using the Beer-Lambert law:

$$C = \frac{OD_{260}}{L} \times \varepsilon \times D$$

Where, C is the concentration of nucleic acid sample in ng/µl
OD_{260} is the blank-normalized absorbance of the nucleic acid sample at 260 nm
L is the path length in cm
ε is the extinction coefficient depending on the type of nucleic acid measured (50 µg/ml for double stranded DNA, 33 µg/ml for single stranded DNA and 40 µg/ml for RNA)
D is the dilution factor

Because proteins and some reagents commonly used in nucleic acids isolation have absorbance at 260 nm, this can lead to increase the measured concentrations than the true values. However, unlike nucleic acids, most of those contaminants have absorbance at 280 nm. Therefore, the ratio of OD_{260}/OD_{280} can be used to control DNA or RNA purity. Generally, values of OD_{260}/OD_{280} ratio should be ≥ 1.8 for DNA and of about 2.0 in case of RNA. When a ratio less than 1.8 is obtained, it indicates a contaminated DNA or RNA sample. The most common sources of contamination include proteins, phenols, guanidine and other aromatic compounds that may be invplved in the purification process.

2.3. Enzymatic modifications of DNA

2.3.1. Restriction digestion

Type II restriction endonucleases were utilized to cut DNA inserts and vectors in a manner that allows subsequent ligation between the fragments cut with the same (or compatible) restriction enzymes. Typically, each type II restriction enzyme recognizes a specific palindromic sequence of 4–8 bp and cuts the DNA double strands within or close to the recognition sequence generating 3′ or 5′ overhangs suitable for subsequent sticky-end ligation. Some restriction enzymes do not generate overhangs and their products are suitable for blunt end ligation. However, in the course of this work, only enzymes generating sticky ends were used to construct expression plasmids. The restriction enzymes were chosen so that their recognition sequences were present in the

multiple cloning site (MCS) of the vector but not in the insert sequence. The composition of the digestion reactions were as follows:

Component	Inserts		Vectors		Constructed plasmids	
DNA solution	add to 10.0	µl	add to 20.0	µl	2.0	µl
10X buffer*	1.0	µl	2.0	µl	2.0	µl
1st restriction enzyme*	0.5–1.0	µl	0.5–1.0	µl	0.5–1.0	µl
2nd restriction enzyme*	0.5–2.0	µl	0.5–2.0	µl	0.5–2.0	µl
dH_2O	-		-		add to 20.0	µl

* During this work, the following combinations were used:
*Eco*RI/*Pac*I: were used in buffer *Eco*RI, 0.5 µl *Eco*RI, 2.0 µl *Pac*I
*Spe*I (*Bcu*I)/*Pac*I: were used in buffer PacI, 1.0 µl *Spe*I (*Bcu*I), 0.5 µl *Pac*I
*Bam*HI/*Hin*dIII: were used in buffer *Bam*HI, 0.5 µl *Bam*HI, 1.0 µl *Hin*dIII
The vector and insert digestions were incubated for 3 h at 37 °C, while plasmid digestion was incubated for 1 h at 37 °C.
Restriction digestion followed by agarose gel electrophoresis was used to control for positive insert ligation in the constructed plasmids. In addition, *Dpn*I, a type IIM restriction endonuclease, was used to digest the methylated template plasmids during site-directed mutagenesis protocol (III.2.10).

2.3.2. Ligation

DNA ligation is the technique by which two DNA fragments (for example, linearized vector and insert) are connected together through the formation of a phosphodiester bond between a 5′ phosphate group in one fragment and a free 3′ hydroxyl group in the other. This reaction is catalyzed by T4 DNA ligase and requires ATP and Mg^{2+} as cofactors. This enzyme can be used in blunt or sticky end ligations and can also repair single-strand nicks in duplex DNA, RNA or DNA/RNA hybrids.
The vectors used were either pGEM-T easy vector for sequencing of the RACE products or pESC-URA and pESC-TRP vectors for expression in yeast cells.
The components of the ligation reaction were as follows:

Insert DNA	6	µl
Linearized vector DNA	2	µl
10X T4 DNA ligase buffer	1	µl
T4 DNA ligase	0.2	µl
50 % PEG 4000	1	µl

The used insert to vector molar ratios were in the range of 1:1 to 5:1, the ligation reaction was incubated overnight at 4 °C.
In the course of this work an instant Sticky-end Ligase Master Mix from NEB was also utilized. The components of the reaction were as follows:

Insert DNA	3.5	µl
Linearized vector DNA	1.5	µl
2X instant ligase master mix	5	µl

The components were mixed thoroughly by pipetting 7–10 times, incubated for 10 min at room temperature. Subsequently, 5 µl of the transformation mixture were transformed in DH5α *E. coli* as described in (III.3.3.1).

2.3.3. Dephosphorylation of the digested vectors

During the ligation step, the digested vector can be recircularized by ligation with its digested fragment instead of the desired insert, resulting in an empty vector that carries the same marker as the desired plasmid. Both will be processed together and could only be distinguished after plasmid isolation and control of the presence of the right insert by restriction digestion or PCR. Dephosphorylation of the vector after digestion serves to the remove the 5′ phosphate groups from the vector as well as from the cut fragment to minimize the possibility of self-ligation. On the other hand, the digested insert (which should not be dephosphorylated) has the 5′ terminal phosphates and can be ligated efficiently to the dephosphorylated plasmid giving a circular plasmid with nicks (open circular plasmids) that could be transformed into competent *E. coli* and the nicks are removed by the DNA-repair mechanism in the bacteria. Therefore, most of the transformed bacterial cells will contain the recombinant but not the recircularized plasmid.
FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) was used for the dephosphorylation of the digested vectors. The reaction components were as follows:

Digested vector	3–4	µg
10X reaction buffer for AP	2	µl
FastAP Thermosensitive Alkaline Phosphatase	2	µl
dH_2O	add to 20	µl

The reaction was incubated for 10 min at 37 °C followed by heat-inactivation of the enzyme for 5 min at 75 °C. The DNA is purified from the protein and salts in the reaction mixture as indicated in (III.2.1.5).

2.4. Agarose gel electrophoresis

Native agarose gel electrophoresis is the method used to separate DNA fragments according to their sizes in a cross-linked agarose matrix by applying an electric field across the gel matrix. As the phosphate groups on the DNA backbone are negatively charged, the DNA will migrate to the positive pole. The migration rate depends on the ionic strength of the used buffer, the size and shape of DNA and the concentration of agarose in the gel. Normally, 1 % agarose gel can be used to separate DNA fragments from 0.2–20 kb. However, the concentration can be increased to 2 % for a better resolution of lower-size fragments (less than 500 bp) or decreased to 0.5 % for a better resolution of high-size fragments. Electrophoresis is mainly a sieving process; where large fragments are entangled in the matrix and, therefore, run more slowly than small fragments. A linear relation exists between the logarithm of the fragment length and its relative migration distance on the gel.
To prepare a medium-size 1 % agarose gel enough for 15–20 samples, 80 ml of 1X TAE buffer (II.8.4) were added to 800 mg of agarose and the mixture was cooked in the microwave for about

1:30 min until it became a transparent solution. When the solution cooled down to about 60 °C, 3 µl of Midori Green were added, thoroughly mixed and directly poured in a gel tray to which a suitable comb is attached. When the gel solidified, the comb was removed, the tray was transferred to the electrophoresis chamber, loading dye (II.8.4) was added to the samples (at 1X final concentration), briefly mixed by pipetting and loaded into the wells formed by the comb. The gel was run in TAE buffer for 30–35 min at 120 V and 400 mA. 5 µl of DNA ladder were loaded into one of the wells to help identifying the sizes of the detected bands. The gel was visualized with a UV-equipped transilluminator. The desired band(s) were cut and DNA was isolated from them as described in (III.2.1.4)

2.5. Primer design

The following points should be taken into consideration during primer design:

To bind specifically to the desired template, the ideal length of the designed primers should be between 18–24 bp, with a GC content in the 40–60 % range and melting temperatures (T_m) ranging from 55 to 65 °C. The selection of the T_m of the primer depends on the application in which it would be used. For example, when a large mix of DNA is used as a template like in the case of genomic DNA or cDNA templates, a primer with relatively higher T_m (typically > 60 °C) should be designed to increase specificity and allow the use of touchdown PCR (III.2.7.2). While in case of using plasmid DNA as a template as in the sequencing reactions, a primer with a T_m of about 50 °C would be sufficient. The melting temperatures for the forward and reverse primers should not differ by more than 5 °C. It is also advisable that the 3′-end nucleotide of the primer should be a C or a G. This GC clamp helps to anchor the primer on its template for efficient priming. T_m above 72 °C as well as runs of multiple Cs or Gs should be avoided especially at the 3′-end as they can lead to mispriming.

Whenever possible, significant homology to other regions on the desired template or to other templates in the cDNA pool should be avoided. This could be checked by performing a BLAST search against the available datasets. Palindromic sequences at the 3′-end of the primes should be also avoided as the can cause the formation of homo- or heterodimers. The tendency of a primer to make hairpins or primer dimers could be checked online (IDT Oligoanalyzer: http://eu.idtdna.com/calc/analyzer).

When designing primers with restriction sites at their 5′-ends for subsequent cloning of the amplified fragment, 3–4 extra nucleotides should be added before the restriction sites to help the enzyme cut the linear amplified DNA. The added restriction sites and extra nucleotides should not be taken into consideration when calculating the T_m.

2.6. Rapid amplification of cDNA ends (RACE)

Rapid amplification of cDNA ends (RACE) is a technique used to recover full-length sequence of cDNA clones using a known part of their internal sequences. The process is done in two halves (3′-RACE and 5′-RACE) and involves two steps. In the first step, a single-stranded cDNA is synthesized by primer extension using reverse transcriptase with the introduction of a 3′- or 5′-anchor. In the second step, PCR is used to amplify the synthesized cDNA using a gene-specific

primer (GSPs) designed from the known internal sequence together with a primer similar to the anchor sequence. The amplification product is subsequently cloned into a suitable vector and sequenced to identify the missing part.

2.6.1. 3′-RACE

In the 3′-reaction, an oligo(dT) primer with a unique anchor sequence at its 5′-end primes the synthesis of a DNA copy of the mRNA starting from the poly(A) tail by the action of reverse transcriptase. Subsequently, a PCR amplifies the fragment between a forward GSP from the known internal sequence and a reverse primer similar to the sequence of the introduced anchor. The PCR will amplify from the poly(A) tail to the middle of the gene (**Figure III-1**).

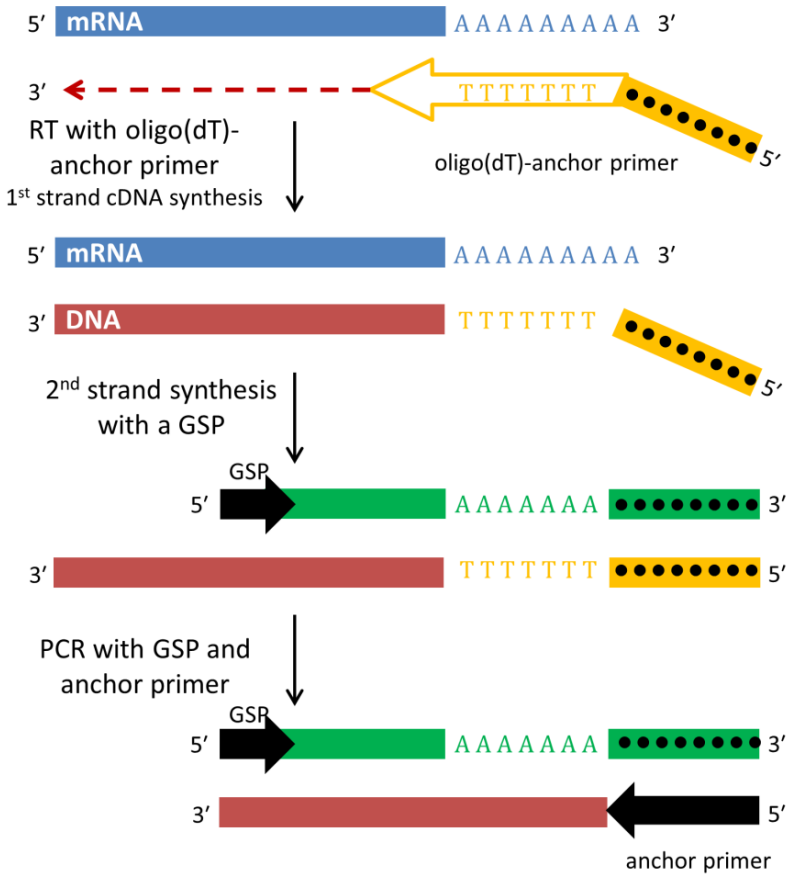


Figure III-1. Schematic presentation of 3′-RACE PCR.

The first-strand cDNA for 3′-RACE was synthesized as follows:
In a nuclease-free PCR tube placed on ice, 1 µl of 3′-RACE CDS primer (10 µM) were added to 5 µg of the total RNA isolated under (III.2.1.1) and the volume was completed to 12.5 µl with DEPC-treated water. After brief mixing, the solution was incubated for 5 min at 65 °C and directly chilled on ice. 7.5 µl of following master mix were then added:

5X reaction buffer	4.0	µl
RiboLock™ RNase inhibitor (40 u/µl)	0.5	µl
dNTPs (10 mM each)	2.0	µl
RevertAid™ H Minus reverse transcriptase (200 u/µl)	1.0	µl
Total volume	**7.5**	**µl**

The reaction mixture was incubated for 90 min at 42 °C, followed by heating for 10 min at 70 °C to deactivate the enzyme. The synthesized 3′-RACE ready cDNA was utilized as a template for PCR as described under (III.2.7).

2.6.2. 5′-RACE using TdT

In the 5′-reaction, a reverse GSP from the internal known sequence is used to initiate DNA synthesis from mRNA by reverse transcriptase. Subsequently, an artificial poly(A) tail is added to the 3′ end of the synthesized DNA strand by the action of terminal deoxynucleotidyl transferase (TdT) in the presence of dATP. An oligo(dT) with a 5′-anchor sequence anneals to the introduced adenosines. Then, A primer with the same sequence as the anchor together with a reverse GSP are used in the PCR amplification for the 5′-reaction (**Figure III-2**).

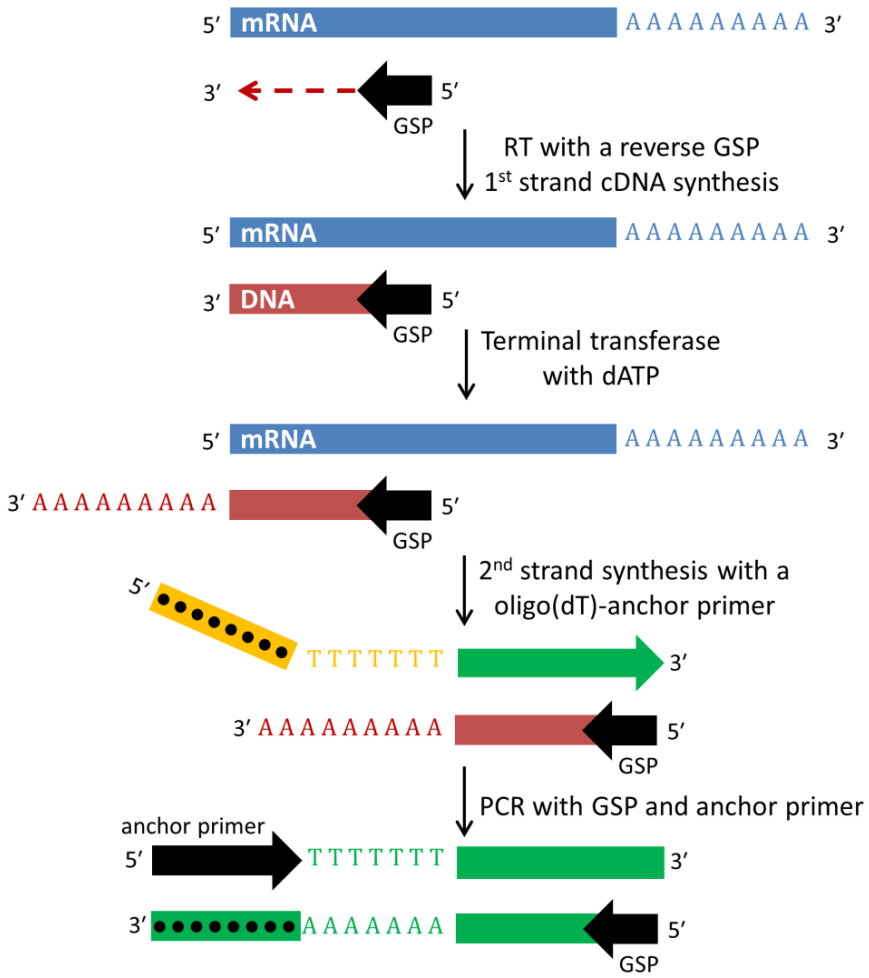


Figure III-2. Schematic presentation of 5′-RACE PCR using TdT.

2.6.3. 5′-RACE using SMART system

SMART 5′-RACE (standing for: Switching Mechanism At 5′-end of RNA-Transcript) is almost similar to TdT 5′-RACE except that the 5′-anchor is introduced during the reverse transcription reation. SMARTscribe™ reverse transcriptase adds 3–5 cytosine nucleotides to the 3′-end of the synthesized cDNA strand because of its terminal nucleotide transferase activity. Next, SMART II™ oligo anneals to the introduced cytosines and template switching occurs allowing the oligo to serve as a template for the SMARTScribe™ RT, thus, generating a complete cDNA copy of the original RNA with an anchor attached at its end. As in TdT 5′-RACE, a forward primer having the same sequence as the introduced 5′-anchor together with a reverse GSP are used in the PCR amplification (**Figure III-3**).

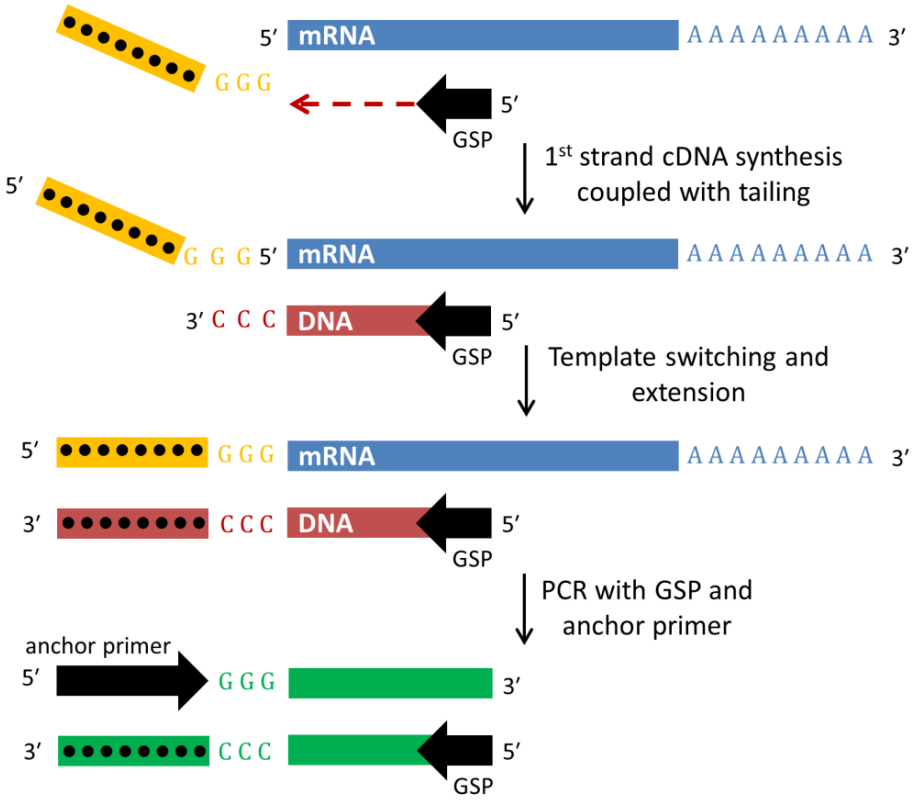


Figure III-3. Schematic presentation of 5′-RACE using SMART system.

The first-strand cDNA for 5′-RACE was synthesized by this method as follows:
In a nuclease-free PCR tube placed on ice, 1 µg of the total RNA (III.2.1.1) were mixed with 1 µl of a reverse GSP (10 µM) and the volume was completed to 3.75 µl with DEPC-treated water. The mixture was incubated for 3 min at 72 °C followed by 2 min at 42 °C, centrifuged for 10 s at 13000 rpm. 6.25 µl of the following master mix were then added:

SMART II™ A oligo (10 µM)	1.0	µl
5X reaction buffer	2.0	µl
DTT (20 mM)	1.0	µl
dNTPs (10 mM each)	1.0	µl
RiboLock™ RNase inhibitor (40 u/µl)	0.25	µl
SMARTScribe™ reverse transcriptase (100 u/µl)	1.0	µl
Total volume	**6.25**	**µl**

The reaction mixture was incubated for 90 min at 42 °C, followed by heating for 10 min at 70 °C to deactivate the enzyme. The synthesized 5′-RACE ready cDNA was utilized as a template for PCR as described under (III.2.7).

2.7. Polymerase chain reaction (PCR)

It is a technique developed by Kary Mullis in the 1980s in which billions of copies of a specific DNA sequence are synthesized by repeated reciprocal extension of two primers oppositely hybridized to the separate strands of a particular DNA sequence by the action of a DNA polymerase (Mullis, 1994). The reaction requires a DNA template (genomic, plasmid, or cDNA), a pair of primers that hybridize in opposite directions to the peripheries of the target sequence, deoxynucleotide triphosphates (dNTPs), and a thermostable DNA polymerase. DNA polymerase synthesizes complementary strands of DNA using a DNA template and primers by adding single nucleotides to the 3′-end of the primer. The process is repeated and the synthesized DNA is used as a template resulting in exponential amplification of the target sequence. The reaction proceeds in three main steps that are repeated in 25–35 cycles:

The denaturation step: heating the DNA for 20–30 s at 95–98 °C leads to disruption of the hydrogen bonds between the complementary bases resulting in the separation of the double strands to form single stranded DNA.

The annealing step: the temperature is decreased to 45–70 allowing the primers to anneal to their specific template. The annealing temperature used usually is 3–5 °C below the T_m of the primers, this prevents non-specific annealing of the primers to other template DNA molecules present in the reaction mixture.

The extension step: the temperature is raised again to 72 °C which is the optimum temperature for most thermostable DNA polymerases. The duration of this step depends on the length of the fragment to be amplified and the speed by which the enzyme adds nucleotides to the newly-synthesized strand.

The PCR product is analyzed by agarose gel electrophoresis (III.2.4) and could be separated as previously described (III.2.1.4).

2.7.1. Standard PCR

This method was used during this work to check for proper ligation of the insert in the constructed plasmids and occasionally in RACE-PCR. The components of the reaction were as follows:

Component			Final concentration	
10X reaction buffer	2.5	µl	1	X
dNTPs (10 mM each)	1	µl	400	µM each
Forward primer (10 µM or pmol/µl)	1	µl	0.4	µM
Reverse primer (10 µM or pmol/µl)	1	µl	0.4	µM
Template DNA	1	µl	1–10	ng/µl
Taq DNA Polymerase (5 U/µl)	0.25	µl	1.25	U/µl
dH_2O	add to 25	µl		

The cycling was performed as follows:

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	1
Denaturation	95 °C	30 s	35
Annealing	T_m - 5 °C	30 s	
Extension	72 °C	2 min 30 s	
Final extension	72 °C	8 min 30 s	1
Storage	12 °C	Pause	

2.7.2. Touchdown PCR

This method was frequently used during this work in RACE-PCR to increase specificity and decrease the possibility of non-specific amplifications. To use touchdown PCR, primers were designed with T_m around 65 °C. In the first cycle, a high annealing temperature was used that was equal to the lowest T_m of the used primer pair. The annealing temperature was decreased by -0.5 °C per cycle for the first 10 cycles, followed by 30 cycles with an annealing temperature of T_m - 5. This allows enrichment of the target sequence during the first few cycles due to the high annealing temperatures used, followed by exponential amplification in the next cycles which would surpass that of any non-specific sequence due to the initial enrichment. The reaction components were the same as in standard PCR but the cycling was performed as follows:

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	1
Denaturation	95 °C	30 s	10
Annealing	T_m (ΔT = -0.5 °C/cycle)	30 s	
Extension	72 °C	2 min 30 s	
Denaturation	95 °C	30 s	30
Annealing	T_m - 5 °C	30 s	
Extension	72 °C	2 min 30 s	
Final extension	72 °C	8 min 30 s	1
Storage	12 °C	Pause	

2.7.3. Nested PCR

This is another technique that could be used to increase the amplification specificity, whereby, the PCR product of one reaction is used as a template for the subsequent PCR using one or both primers in the second amplification to the 3′-direction of those originally used in the first reaction. This technique can be used with either the standard or touchdown PCR.

Occasionally, an enhancer solution (II.8.7) was used in *Taq*-catalyzed PCR to increase amplification efficiency by decreasing the DNA melting temperature and therefore decrease the possibility of the formation of secondary structures such as hairpins which would interfere with the polymerase function (Ralser et al., 2006).

2.7.4. PCR using high-fidelity DNA polymerases

For experiments requiring absolutely correct DNA sequences such as cloning of genes into expression vectors and site-directed mutagenesis, high-fidelity (proofreading) DNA polymerases have been employed. The advantage of those polymerases is that they have very low replication error rate compared to *Taq* DNA polymerase. This is due to their ability to bind preferentially to the correct nucleotide during polymerization. However, if a mismatched nucleotide is inserted, those enzymes possess an extra line of defense, their 3′→5′ exonuclease domain, also known as the proofreading domain, enables the polymerase to excise the incorrect nucleotide in the 3′→5′ direction, thus increasing fidelity of replication. Unlike the products of *Taq*-polymerase, products of high-fidelity polymerases are blunt-ended and cannot be used directly for TA cloning.

The components of the PCR were as follows:

Component					**Final concentration**	
5X Phusion HF buffer *	10	µl	4	µl	1	X
dNTPs (10 mM each)	1	µl	0.4	µl	200	µM each
Forward primer (10 µM or pmol/µl)	2.5	µl	1	µl	0.5	µM
Reverse primer (10 µM or pmol/µl)	2.5	µl	1	µl	0.5	µM
Template cDNA	1	µl	1	µl	50 - 250	ng
Phusion Hot Start II DNA Polymerase (2 U/µl) *	0.5	µl	0.2	µl	0.02	U/µl
dH_2O	add to 50	µl	add to 20	µl		

* Q5 Hot Start High-Fidelity DNA Polymerase (NEB) was also successfully used with its own reaction buffer.

The cycling was performed as follows:

Step	**Temperature**	**Time**	**Cycles**
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	35
Annealing	T_m + 3 °C	30 s	
Extension	72 °C	1 min 30 s	
Final extension	72 °C	10 min	1
Storage	12 °C	Pause	

2.8. Gene transcript analysis by real-time qRT-PCR

Total RNA was extracted from control and yeast-extract-treated *H. calycinum* cell cultures at 4, 8, 12, 16, 20, 24, 36 and 48 h post-elicitation as described in (III.2.1.1.1). On-column digestion was applied to each sample using the RNase-Free DNase Set (Qiagen) to get rid of genomic DNA contamination. The RNA quality was checked on a gel. Concentrations and 260/280 ratios were determined using the SimpliNano™ Spectrophotometer (GE Lifesciences). For each time point, cDNA was synthesized from 1 µg RNA using iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad). Subsequent measurements were performed on the CFX Connect™ Real-Time PCR Detection System (Bio-Rad) using iTaq™ Universal SYBR® Green Supermix (Bio-Rad), following the manufacturer's protocol. The 20 µl reaction contained 1 µl cDNA, 10 µl (2x) supermix and 0.5 µM of each primer (II.3.5). Samples were initially denatured at 95 °C for 30 s, run for 40 cycles at 95 °C for 5 s and 59 °C for 30 s. Data were recorded after the annealing/extension step. The specificity of the amplification product was confirmed from melt curves and by running the samples on an agarose gel. Pooled cDNA of all time points was used in a serial dilution to determine the efficiency of amplification for each primer pair. Based on the Cq values obtained during the efficiency tests, 1 µl of 1:50 dilution of the original cDNA of each time point was used as a template for the subsequent expression analysis. *Actin* and *Histone H2A* served as reference genes. The non-elicited sample (0 h) served as a calibrator. Amplification efficiencies as well as normalized expression relative to the calibrator were determined using the Bio-Rad CFX Manager software (version 3.1, Bio-Rad).

2.9. C-terminal exchanges

Reciprocal exchange of the terminal 55 amino acids between HpCYP81AA1 and HpCYP81AA2 was performed using the one-pot fusion PCR protocol described in (Buttner and Barleben, 2012). In this protocol, a PCR is set up with two templates, a small concentration of an overlapping primer between the two domains to be fused and higher concentrations of the 5′- and 3′-gene specific primers (GSPs) with suitable restriction sites for subsequent cloning. In the first few PCR cycles, the overlapping primer (green/red arrow) together with the 3′-GSP (red arrow) generate a mega-primer (red/green arrow) from the C-terminal domain. After depletion of the overlapping primer, the synthesized mega-primer together with the 5′-GSP (green arrow) can prime the formation of the chimeric gene from the N-terminal domain. The formed chimeric gene is exponentially amplified with the 5′- and 3′-GSPs (**Figure III-4**).

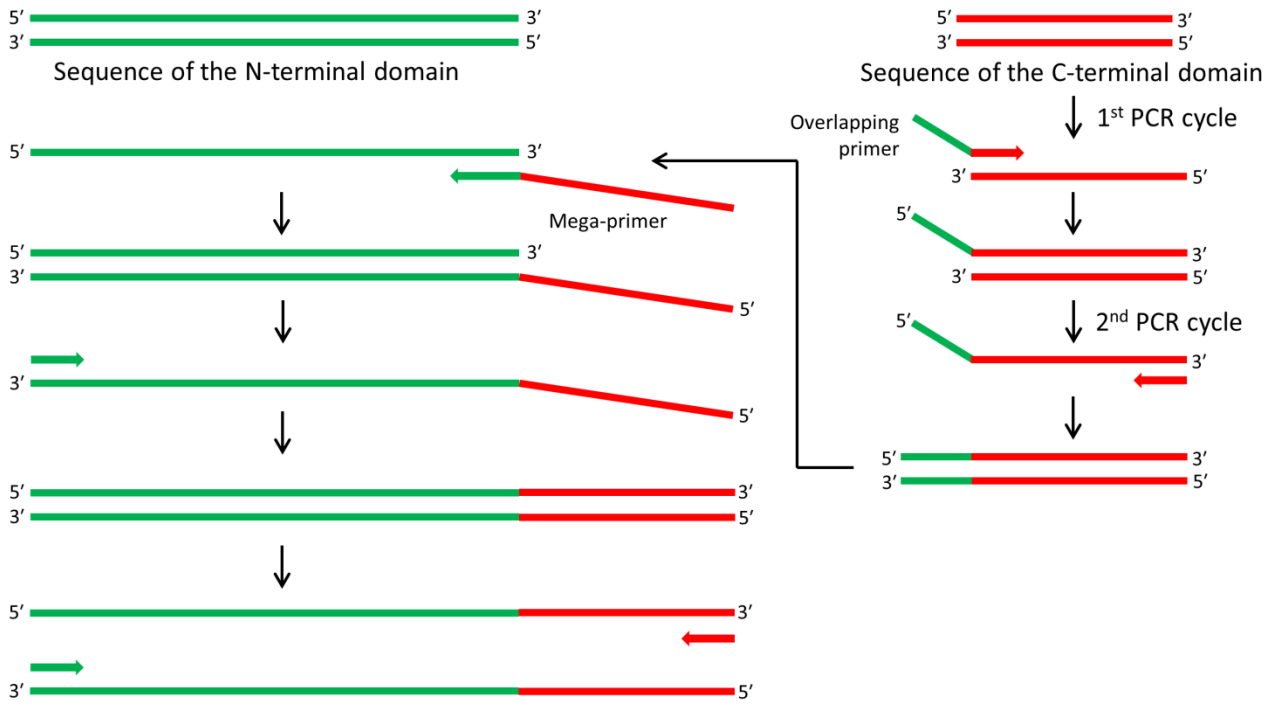


Figure III-4. Schematic presentation of the one-pot fusion PCR protocol; adapted from Buttner and Barleben (2012).

The components of the PCR were as follows:

Component			Final concentration	
5X Phusion HF buffer	10	µl	1	X
dNTPs (10 mM each)	1	µl	200	µM each
5′-GSP (10 µM)	1	µl	0.2	µM
3′-GSP (10 µM)	1	µl	0.2	µM
Overlapping primer (1 µM)	2	µl	0.04	µM
Template plasmid 1 (1 µg/µl)	0.5	µl	10	ng/µl
Template plasmid 2 (1 µg/µl)	0.5	µl	10	ng/µl
Phusion Hot Start II DNA Polymerase (2 U/µl)	0.5	µl	0.02	U/µl
dH_2O	add to 50	µl		

HpCYP81AA1 with the C-terminus of HpCYP81AA2 (HpCYP81AA1-AA2) was generated using Hp81AA1-*Spe*I-F, Hp81AA2-*Pac*I-R and AA1-AA2-ex primers. While CYP81AA2 with the C-terminus of CYP81AA1 (HpCYP81AA2-AA1) was generated using Hp81AA2-*Spe*I-F, Hp81AA1-*Pac*I-R, and AA2-AA1-ex primers.

The PCR started with an initial denaturation at 98°C for 30 s, then 35 cycles at 98°C for 10 s, 60°C for 30 s, and 72°C for 1:30 min, followed by a final extension at 72°C for 10 min. The PCR product with the expected size (~1500 bp) was excised from agarose gel (III.2.5), purified (III.2.1.4), digested with *Spe*I and *Pac*I restriction enzymes (III.2.3.1), ligated to pESC-URA-HcCPR2 plasmid previously digested with the same restriction enzymes (III.2.3.2), transformed into DH5α (III.3.3.1) and sequenced to verify the generation of a chimeric gene.

2.10. Site-directed mutagenesis (SDM)

Site-directed mutagenesis (SDM) was performed according to the QuickChange™ protocol following the modified primer design described by Liu and Naismith (2008). In the original design, the mutagenic primers anneal to the same sequence on opposite strands of the parental plasmid to replicate its sequence and introduce the desired mutations. Therefore, the product has a strand break (nick) and cannot be used as a template for the subsequent PCR cycles (**Figure III-5a**). In addition, the primers are completely complementary to one another, and therefore, they tend to form primer-dimers instead of annealing to the template plasmid DNA, resulting in primers depletion and finally low yield of the mutated product.

The modification suggested by Liu and Naismith (2008) utilizes partially complementary primers with a short overlapping part and a significantly longer non-overlapping parts (**Figure III-6**) to decrease the possibility of primer-dimer formation. Moreover, the non-overlapping parts of the primers are long enough to bridge the nicks (**Figure III-5b**) and therefore can use the parental plasmid as well as newly-synthesized DNA as templates resulting in increasing the PCR efficiency and reducing the amount of template plasmid needed.

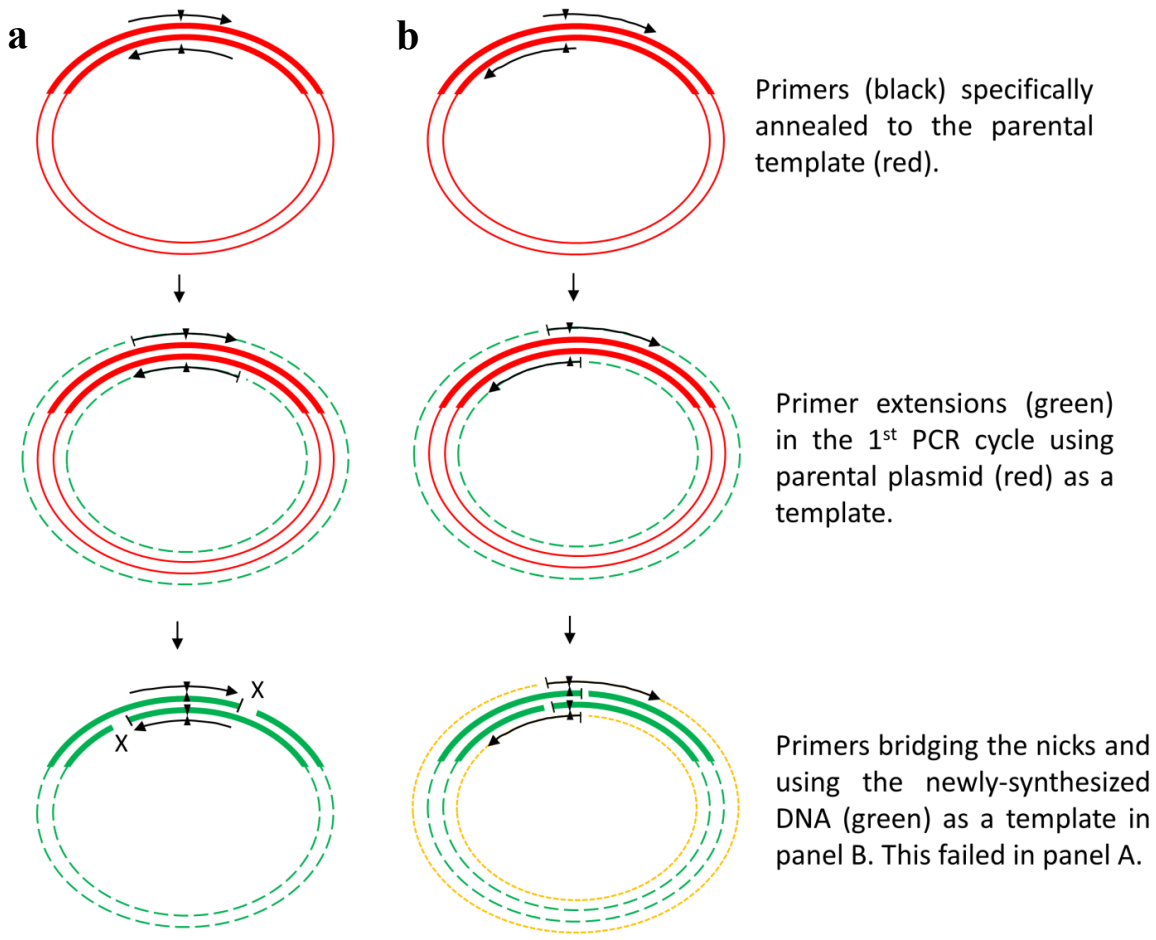


Figure III-5. Schematic presentation of SDM-PCR. (**a**) Using a pair of complementary primers as recommended by the QuickChange™ protocol, and (**b**) using partially overlapping primers that gap the nicks in the newly-synthesized plasmids and use them as templates; adapted from (Liu and Naismith, 2008).

The following points should be considered when designing primers for SDM (**Figure III-6**)

- Each primer pair consist of long non-overlapping parts of about 26 nucleotides at their 3′ ends and an overlapping (complementary) part of about 13 nucleotides at their 5′ ends.
- The melting temperature of the non-overlapping parts should be 5–10 °C higher than the melting temperature of the complementary part.
- Single or multiple mutations could be introduced either in the complementary or non-overlapping parts.

This primer design can also be used for the introduction of insertion or deletion mutations as illustrated in (**Figure III-6**)

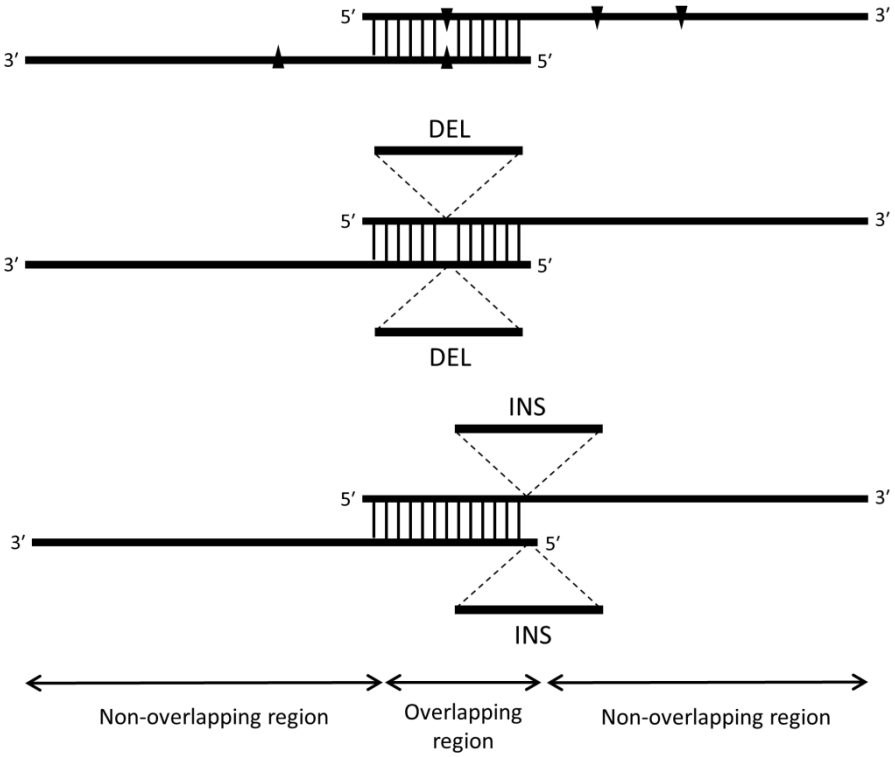


Figure III-6. Schematic presentation of the primer design for SDM, deletion (DEL) and insertion (INS) mutations; adapted from (Liu and Naismith, 2008).

The components of the PCR were as follows:

Component			**Final concentration**	
5X Phusion HF buffer *	10	µl	1	X
dNTPs (10 mM each)	1	µl	200	µM each
Forward primer (10 µM or pmol/µl)	2.5	µl	0.5	µM
Reverse primer (10 µM or pmol/µl)	2.5	µl	0.5	µM
Template plasmid (1:20 dilution)	1	µl	0.5–1.0	ng/µl
Phusion Hot Start II DNA Polymerase (2 U/µl) *	0.5	µl	0.02	U/µl
dH_2O	add to 50	µl		

* The SDM-PCR was sometimes successfully performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB) using its own reaction buffer.

The cycling was performed as follows:

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	25
Annealing	70 °C	30 s	
Extension	72 °C	6 min 30 s	
Final extension	72 °C	10 min	1
Storage	12 °C	Pause	

1 µl of *Dpn*I (10 U/µl) were added to the PCR product and incubated at 37°C for 3 h to digest the methylated parental plasmid. 5 µl of the digestion mixture were run on an agarose gel (III.2.4) to control for the presence of a band at the expected size of the mutated plasmid; another 5 µl were directly transformed into DH5α *E. coli* (III.3.3.1), plasmids were isolated (III.2.1.2) and sequenced to confirm the introduction of the designed mutations. The sequences of the primers used for SDM are listed in (II.3.3); the templates and primers combinations used to generate each mutant were as follows:

Template	Primers	Mutant
Wild-type CYP81AA1	AA1-A373S-F AA1-A373S-R	A373S
AA1-AA2	AA1-A373S-F AA1-A373S-R	AA1-AA2 A373S
A373S	AA1-AS-SL-F AA1-A373S-R	A373S/S376L
AA1-AA2 A373S	AA1-AS-SL-F AA1-A373S-R	AA1-AA2 A373S/S376L
A373S/S376L	AA1-T482A-F AA1-T482A-R	A373S/S376L/T482A
A373S/S376L	AA1-TALMRKKN-F AA1-T482A-R	A373S/S376L/T482A/L485M/R487K/ K488N
A373S/S376L	AA1-T482V-F AA1-T482uni-R	A373S/S376L/T482V
A373S/S376L	AA1-T482F-F AA1-T482uni-R	A373S/S376L/T482F
A373S/S376L/T482A	AA1-W113F-F AA1-W113F-R	W113F/A373S/S376L/T482A
CYP81AA1 A373S/S376L/T482A	AA1-I220P-F AA1-I220P-R	I220P/A373S/S376L/T482A
A373S/S376L/T482A/L485M/R487K/ K488N	AA1-VIMT-F AA1-VIMT-R	A373S/S376L/T482A/L485M/R487K/ K488N/V302I/ M303T
A373S/S376L/T482A/L485M/R487K/ K488N/ V302I/M303T	AA1-W113F-F AA1-W113F-R	W113F/A373S/S376L/T482A/L485M/ R487K/K488N/V302I/M303T
W113F/A373S/S376L/T482A/L485M/ R487K/K488N/V302I/M303T	AA1-I220P-F AA1-I220P-R	W113F/I220P/A373S/S376L/T482A/ L485M/R487K/K488N/V302I/M303T
Wild-type CYP81AA2	AA2-S375A-F AA2-S375A-R	S375A
AA2-AA1	AA2-S375A-F AA2-S375A-R	AA2-AA1 S375A

Continued:

Template	Primers	Mutant
Wild-type CYP81AA2	AA2-L378S-F AA2-L378S-R	L378S
Wild-type CYP81AA2	AA2-A483T-F AA2-A483T-R	A483T
S375A	AA2-SALS-F AA2-S375A-F	S375A/L378S
AA2-AA1 S375A	AA2-SALS-F AA2-S375A-F	AA2-AA1 S375A/L378S
S375A	AA2-A483T-F AA2-A483T-R	S375A/A483T
L378S	AA2-A483T-F AA2-A483T-R	L378S/A483T
S375A/L378S	AA2-A483T-F AA2-A483T-R	S375A/L378S/A483T
S375A/L378S	AA2-ATMLKRNK-F AA2-A483T-R	S375A/L378S/A483T/M486L/K488R/ N489K

3. Bacteria and yeast methods

3.1. Determination of microbial count

The count of *Escherichia coli* or *Saccharomyces cerevisiae* during culture growth was determined using a spectrophotometer by measuring the optical density at 600 nm of 1 ml of bacterial or yeast culture with fresh media as blank. An OD_{600} of 0.1 corresponds 1×10^8 cells/ml in case of *E. coli*, or 3×10^6 cells/ml of *S. cerevisiae* (Lech and Brent, 2001; Treco and Winston, 2001).

3.2. Preparation of competent cells

3.2.1. Preparation of competent *E. coli* cells

The DH5α *E. coli* cells described in (II.1.3) were made competent using the calcium chloride method (Mandel and Higa, 1970). A single *E coli* colony was inoculated in a 5 ml LB tube and incubated overnight at 37 °C with shaking at 200 rpm. 1 ml of the overnight culture was diluted in 50 ml LB medium to obtain an OD_{600} of 0.2–0.3, the culture was allowed to grow until reaching an OD_{600} of 0.6–0.9. The bacteria were centrifuged in a 50 ml sterile falcon tube for 5 min at 4 °C and 5000 rpm. The supernatant was discarded and the pellet was washed with 50 ml 50 mM $CaCl_2$ precooled solution, centrifuged again, the supernatant was discarded and the pellet was resuspended in 20 ml 50 mM $CaCl_2$ solution and kept on ice for 20 min. The suspension was centrifuged again and the pellet was resuspended in 860 µl 50 mM $CaCl_2$ solution, transferred to a sterile 1.5 ml Eppendorf tube and incubated on ice for 15 min. 140 µl of sterile glycerol were added and briefly mixed with the tube content. The suspension was divided into 50 µl aliquots, stored at -80 °C for later transformation as described in (III.3.3.1).

3.2.2. Preparation of competent *Saccharomyces cerevisiae* cells

The INV*Sc*1 *Saccharomyces cerevisiae* cells described in (II.1.3) were made competent using the *S.c.* EasyComp Transformation Kit (Invitrogen), based on the lithium acetate method described in (Dohmen et al., 1991; Becker and Lundblad, 2001). Yeast cells from a stock culture prepared as

described in (III.3.4.2) were streaked onto YPD plate and incubated at 30 °C for 48 h. A single colony was picked and inoculated in 10 ml YPD medium, incubated at 30 °C and 250 rpm for about 24–28 h until the OD_{600} reached 2.6–5.0. About 600 µl of the starter culture were diluted in 10 ml fresh YPD medium to obtain an OD_{600} of 0.2–0.4. The cells were grown for about 6 h until the OD_{600} reached 0.6–1.0 (it is advisable to allow the OD_{600} to triple to obtain higher transformation efficiency). The cells were pelleted by centrifugation at 1500 rpm for 5 min at room temperature, the supernatant was discarded. The resulting pellet was washed with 10 ml of Solution I, centrifuged and the supernatant was discarded. The pellet was resuspended in 1 ml of Solution II and divided into 50 µl aliquots, stored at -80 °C for later transformation as described in (III.3.3.3).

3.3. Transformation of competent cells

3.3.1. Transformation of competent *E. coli* cells

A 50 µl aliquot of DH5α *E. coli* cells made competent as described in (III.3.2.1) were thawed slowly on ice, gently mixed by pipetting with 5–10 µl of the ligation mixture described in (III.2.3.2) or 100 ng of the isolated plasmids described in (III.2.1.2 or III.2.1.3) and incubated for 20 min on ice. The transformation was performed by the heat shock method; the mixture was heated in a 42 °C water bath for 45 s followed by rapid cooling on ice for 2 min. Afterwards, 250 µl of SOC medium were added and the transformation mixture was incubated for 60 min at 37 ºC and 200 rpm. 200 µl of the transformation mixture were spread on a suitable LB plate and incubated overnight (16–18 h) at 37 °C.

3.3.2. Control for positive insert ligation

Control for successful ligation of a DNA fragment on pGEM®-T Easy vector could be done by α-complementation of the *lacZ* gene. The *E. coli* strain used (DH5α) carries only the α-galactosidase subunit of the gene and therefore cannot use the substrate X-Gal enzymatically. The multiple cloning site of the vector lies in a β-galactosidase coding region. If no insertion takes place and self-ligation of the vectors occurs, the β-galactosidase will be expressed after induction by IPTG in the medium. Together, the expressed β-galactosidase subunit of the empty plasmid with the α-galactosidase encoded in genome of the bacteria can utilize the substrate X-Gal and convert it into a blue dye. Therefore, bacterial colonies with an empty plasmid appear blue. On the other hand, bacteria containing plasmids, in which the ligation of a DNA fragment in the MCS disrupts the expression of the β-galactosidase are not capable of utilizing X-Gal enzymatically and form white colonies.

3.3.3. Transformation of competent *Saccharomyces cerevisiae* cells

Transformation was performed using *S.c.* EasyComp Transformation Kit (Invitrogen). A 50 µl aliquot of INV*Sc*1 cells made competent as described in (III.3.2.2) were thawed at room temperature. 1–5 µg of the expression plasmids isolated from *E. coli* as described in (III.2.1.2) dissolved in a maximum volume of 5 µl were added to the cells. 500 µl of Solution III equilibrated to room temperature were subsequently added and the transformation mixture was mixed by vortexing vigorously. The transformation mixture was incubated in a water bath for 1 h at 30 °C

with mixing by vigorous vortexing every 15 min. 200 µl were spread on appropriate SD dropout plates lacking the marker(s) in the transformed plasmid(s) for auxotrophic selection.

3.4. Preparation of stock cultures

3.4.1. Preparation of *E. coli* stock cultures

A single colony of *E. coli* growing on LB/ampicillin solid medium was inoculated in 5 ml LB liquid medium supplemented with 10 µl of 100 mg/ml ampicillin solution and grown overnight (16–18 h) at 37 °C and 200 rpm. The stock culture was prepared in screw-capped 2 ml tubes by mixing 750 µl of the overnight culture with 250 µl of sterile LB/glycerol (4:6). The cultures were stored at -80 °C.

3.4.2. Preparation of *S. cerevisiae* stock cultures

A single colony of *S. cerevisiae* growing on a suitable selective solid medium was inoculated in 5 ml of the same selective liquid medium, incubated for 24 h at 30 °C and 250 rpm. The stock culture was prepared in screw-capped 2 ml tubes by mixing 250 µl of the starter culture with 300 µl of YPD medium and 450 µl of sterile glycerol (86 %). The cultures were stored at -80 °C.

4. Protein methods

4.1. Protein expression in yeast

The P450s together with CPR proteins were heterologously expressed in INV*Sc*1 yeast cells according to the high density procedure described by Pompon et al. (1996). A single colony from the transformed yeast obtained under (III.3.3.3) was picked, inoculated in a 5 ml of synthetic dextrose medium (SD) and incubated at 30°C with shaking at 250 rpm for 24 h. 1 ml of the grown culture was transferred to 150 ml YPGE medium, grown for 28–30 h until the OD_{600} reached 1.5–1.7, protein expression was induced by the addition of 16.7 ml of 20% sterile-filtered galactose solution (2% final concentration) and the culture was further incubated for 14–16 h at 30°C with shaking at 250 rpm.

4.2. Isolation of the microsomal fraction

The induced yeast cells obtained as described in (III.4.1) were harvested by centrifugation at 5000 rpm for 5 min; the supernatant was discarded while the pellet was resuspended in 10 ml of TEK buffer. The suspension was transferred into a 15 ml falcon tube and centrifuged at 5000 rpm (2900 g) for 5 min. The supernatant was discarded and the pellet was resuspended in 5 ml of TEK buffer, divided equally into two 15 ml falcon tubes (each contained about 4 ml) and incubated at room temperature for 5 min. The suspensions were centrifuged again, the supernatant was discarded and the pellet in each falcon tube was resuspended in 3 ml of TES buffer. 3 gm of 0.45-mm acid-washed glass beads were added to each tube and shaken in the cold room for 20 min to mechanically break the yeast cells. The cracked cells were centrifuged and the supernatant was collected. The sediment in each tube was resuspended in 2 ml of TES buffer, centrifuged and the supernatant was collected. The last step was repeated once. 20 ml of TES buffer were added to the total collected supernatant (14 ml) and then centrifuged at 15000 rpm for 10 min to sediment nuclei and mitochondria. The pellet was discarded and the supernatant was centrifuged at 100000 g for

60 min. The supernatant was discarded and the pelleted microsomes were resuspended in 1 ml of TEG buffer using a Pasteur pipette and stored at -80°C. All steps for microsomal preparation were performed at 0–4°C. The concentration of the microsomal protein was determined as described in (III.4.3).

4.3. Determination of protein concentration

The determination of protein concentrations was performed according to Bradford (1976). The principle of this method is based on the shift of the absorption maximum of Coomassie Brilliant Blue G-250 dye from 465 to 595 nm upon binding to proteins.

In 1.5 ml Eppendorf tubes, 2–5 µl of the protein solutions were completed to 100 µl with deionized water. 900 µl of Bradford solution (II.8.6) were added and mixed by flipping, followed by incubation for 5 min at room temperature. As the absorbance could be affected by the pH of the solution as well as the presence of certain salts and detergents, a blank was carried out by mixing 900 µl of Bradford solution with 2–5 µl of the buffer in which the protein was dissolved completed to 100 µl of deionized water. The blank solution was used to zero the spectrophotometer and the absorbance at 595 nm of test solution was subsequently measured. The concentration of the protein was calculated according to the following equation:

$$C = \frac{OD_{595}}{\text{Bradford factor} \times V}$$

Where, C is the concentration of the protein solution in µg/µl

OD_{595} is the blank-normalized absorbance of the protein solution with Bradford reagent at 595 nm

V is the volume in microliters of the protein solution measured

The Bradford factor is calculated from the slope of the calibration curve generated by plotting the absorbance of different concentrations (1–10 µg/ml) of BSA solution in deionized water on the Y axis against the amount of BSA in micrograms on the X axis.

4.4. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE)

It is a technique employed for the separation of proteins based on their sizes in a cross-linked polymer under the influence of an electric field. Under these conditions, the speed by which a protein migrates is directly proportional to its net charge and inversely proportional to its size. The anionic surfactant sodium dodecyl sulfate serves to coat all proteins uniformly with negative charges causing them to unfold so that the proteins migrations are exclusively dependent on their respective molecular weights.

The polyacrylamide gel is divided into two parts differing in their pH and acrylamide percentage. The upper part (the stacking gel) has a pH of 6.8 and 5% acrylamide, while the lower part (the resolving or running gel) has a pH of 8.8 and 12% acrylamide. The reason for using this composition is to squeeze the protein bands into a very thin layer in the stacking gel before being separated in the resolving gel so that the separation will start at the same time for all samples. The main current-carrying species in SDS-PAGE are the chloride ions and the glycine molecules present in the electrophoresis buffer. The charge of the chloride ions is independent on the pH, whereas, glycine is present in a zwitterionic form at pH 6.8 while it is negatively charged at pH 8.8.

In the stacking gel with a pH of 6.8, the neutral glycine moves slowly behind chloride ions and the proteins creating a region of low conductivity (high voltage drop) causing the proteins to move slowly and therefore deposit as a very narrow band on top of the resolving gel. In the resolving gel with a pH of 8.8, both the glycine and chloride ions are negatively charged moving faster than the proteins which are then separated in the highly cross-linked gel according to their sizes.
Acrylamide chains are cross linked with bisacrylamide by the action of the radical donor APS in the presence of TEMED which acts as a radical stabilizer. To prepare the gels, the components of resolving gel were mixed together in a flask, poured between the glass plates of the instrument and covered with a thin layer of water to avoid interaction with air. After gel solidification, the water was removed with a filter paper. The components of the stacking gel were then mixed and poured over the solidified resolving gel, a comb was added to create the sample pockets. The gel was allowed to solidify. The protein samples were denatured in presence of the sample buffer at 95 °C for 5 min. Samples were loaded on the gel and run at 200 V. The run was ended when the dye bromophenol blue reached the front of the gel. The gel was stained by immersing in the staining solution for 30 min with gentle shaking, then destained by immersing in the destaining solution overnight.

4.5. Enzyme assays

4.5.1. NADPH-dependent Cytochrome P450 reductase (CPR) activity assay

The test depends on the ability of NADPH-dependent cytochrome P450 reductase to reduce cytochrome c. The reduced cytochrome c has a characteristic absorption maximum at 550 nm and can be measured spectrophotometrically (Guengerich et al., 2009).
The components of the reaction were as follows:

Phosphate buffer (1 M, pH 7.7)	300	µl
Cytochrome c (1 mM, in 10 mM phosphate buffer)	40	µl
Microsomal protein	5	µg
NADPH (4 mM, in dH_2O)	25	µl
dH_2O	add to 1	ml

The reaction was started by the addition of NADPH and the absorbance is measured at 550 nm. The specific activity of the CPR (µmol min^{-1} mg^{-1}) could be calculated from the following equation:

$$E = (\Delta_{A550} \times V) / (\varepsilon \times t \times p)$$

Δ_{A550} : Change in the absorption at 550 nm
ε : Absorption coefficient cytochrome c (ε = 21000 cm^{-1} mM^{-1})
V : Reaction volume
t : Reaction time
p : Protein amount

4.5.2. Cytochrome P450 activity assay

The standard *in vitro* cytochrome P450 activity assay consisted of the following:

Phosphate buffer (1 M, pH 7.0)	20	µl
Substrate (10 mM, in methanol)	4	µl
Microsomal protein	50	µg
NADPH (4 mM, in dH_2O)	50	µl
dH_2O	add to 200	µl

All ingredients except for the NADPH were mixed, briefly centrifuged and the reaction was started by the addition of 50 µl of 4 mM NADPH solution. The reaction was incubated in a water bath for 1 h at 30 °C. The reaction was stopped by the addition of 20 µl of 1 N HCl and was extracted twice, each with 250 µl ethyl acetate. The combined ethyl acetate extract was evaporated to dryness in a rotational vacuum concentrator at 37 °C, redissolved in 60 µl of HPLC-grade methanol, 25 µl of which were analyzed by HPLC-DAD or LC-MS as described in (III.5.1 and III.5.2).

5. Analytical methods

5.1. High performance liquid chromatography (HPLC)

The following gradients were utilized with the VWR-Hitachi system equipped with ZORBAX eclipse plus C18 rapid resolution column. The mobile phases used were double distilled water containing 0.1 % formic acid (A) and HPLC-grade acetonitrile (C). The flow rate was 1 ml/min and the detection was carried out at a wavelength of 260 nm.

Gradient 1		**Compounds analyzed**
Time	**% C**	
0	5	2,4,6-trihydroxybenzophenone
4	5	2,3′,4,6-tetrahydroxybenzophenone
12	30	1,3,7-trihydroxyxanthone
25	34	1,3,5-trihydroxyxanthone
28	95	
33	95	
35	5	
45	5	

Gradient 2		**Compounds analyzed**
Time	**% C**	
0	5	2,4-dihydroxybenzophenone
4	5	2,3′,4-trihydroxybenzophenone
10	30	2,2′,3,4′-tetrahydroxybenzophenone
22	40	2,2′,4,5′-tetrahydroxybenzophenone
26	95	
30	95	
32	5	
42	5	

Gradient 3		Compounds analyzed
Time	% C	
0	5	2,3′,4,4′,6-pentahydroxy-benzophenone (maclurin)
4	5	1,3,6,7-tetrahydroxyxanthone
12	20	1,3,5,6-tetrahydroxyxanthone
25	27	
28	95	
33	95	
35	5	
45	5	

5.2. Liquid chromatography–Mass spectrometry (LC-MS)

Products of 30 enzymatic incubations were purified by HPLC and corresponding peaks were collected as individual fractions. These fractions were directly infused into the mass spectrometer (3200QTrap mass spectrometer; Applied Biosystems/MDS SCIEX, Darmstadt, Germany), equipped with an electrospray ionization (ESI) interface (Turbo V), using the integrated syringe pump of the 3200 QTrap instrument (Syringe; 1000 µL, i.d. 2.3 mm; Hamilton, Nevada, USA) at a flow rate of 10 µL/min. The mass spectrometer was operated in the positive mode with a source voltage and declustering potential of 5.5 kV and 76 V, respectively. Nitrogen gas was used for nebulization, with the curtain gas, gas 1, and gas 2 settings at 10, 14, and 0, respectively. Parameters were optimized for benzophenones using standard 2,4,6- trihydroxybenzophenone and for xanthones using 1,3,7-trihydroxyxanthone in methanol at 10 µg/ml. The molecular ion peaks $[M + H]^+$ of the products were further analyzed by MS/MS experiments in the enhanced product ion mode (EPI) of the instrument using nitrogen gas for collision-induced dissociation at the high-level setting. The collision energy was 30–50 V. Data acquisition and processing were performed using the Analyst software (version 1.4.2; Applied Biosystems/MDS SCIEX).

6. Bioinformatic methods

6.1. Phylogenetic analysis

Phylogenetic analysis of CPR as well as CYP81 sequences was performed using MEGA6 program (Tamura et al., 2013). The sequences were collected from public databases (II.10) together with the deduced amino acid sequences of the products of the cloned genes within this work and multiple sequence alignment was performed using ClustalW (Larkin et al., 2007). The alignment was used to select the best settings using the "Find Best DNA/Protein Models (ML)" tool in MEGA6.

For CYP81s, the evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model (Le and Gascuel, 2008). The tree with the highest log likelihood (-15166.6046) was displayed. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model (Jones et al., 1992). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 1.9675)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 2.7835 % sites). The tree was drawn to scale,

with branch lengths measured in the number of substitutions per site. The analysis involved 45 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 326 positions in the final dataset. For CPRs, the evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-9756.9348) was displayed. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.6957)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 21.5937 % sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 20 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 599 positions in the final dataset. The percentage of trees in which the associated taxa clustered together is (bootstrap values of 1000 replicates) were indicated for each node.

6.2. Homology modeling

The HHpred server (Soding et al., 2005) was used to identify homologous CYP X-ray crystal structures as 3D-templates for comparative modeling utilizing the software Modeler (Sali and Blundell, 1993). The identified homologous template structures were all in the low sequence identity range (between 20-25 % sequence identity) and an alternative homology modeling attempt using YASARA Structure 14.7.17 was used to identify the most suitable template structures by another independent method (Krieger et al., 2003; Krieger and Vriend, 2014). Structures from two matching CYP-families were identified by sequence search using Psi-BLAST (Altschul et al., 1997) and Psi-Pred secondary structure prediction employing position specific patterns (Jones, 1999). The generated homology models were refined using MD-simulations with the YAMBER (Krieger et al., 2009) force field and ranked according to knowledge-based structural descriptors as implemented in YASARA. The best scoring structures were obtained using the X-ray crystal structure from eukaryotic CYP2B4 (Gay et al., 2010) in the closed conformation with the bound inhibitor ticlopidine (PDB: 3kw4) as template. In the alignment, 400 of 508 target residues (78.7 %) are aligned to template residues of HpCYP81AA1. Among these aligned residues, the sequence identity is 30.8 % and the sequence similarity is 48.0 % ('similar' means that the BLOSUM62 score is > 0). A second reference structure showing good scores was human CYP17A1 (steroid 17-*alpha*-hydroxylase) in closed conformation with bound substrate mimetic. In the alignment, 378 of 508 target residues (74.4 %) are aligned to template residues of HpCYP81AA1. Among these aligned residues, the sequence identity was 27.8 % and the sequence similarity was 47.4 % ('similar' means that the BLOSUM62 score is > 0). This model structure based on CYP17A1 showed a different orientation of the C-terminal beta-loop (SRS-6) and the adjacent α-helix containing SRS-2 and it was crystallized in the dimeric form. The final 3D-models of HpCYP81AA1 and HpCYP81AA2 were generated from a hybrid model using the three best matching template structures: 3kw4 (CYP2B4), 3swz (CYP17A1), 3qz1 (steroid 21-hydroxylase) and 2hi4 (CYP1A2).

6.3. Product docking

The model structures of HpCYP81AA1 and HpCYP81AA2 were overlaid by MUSTANG (Konagurthu et al., 2006) with heme and the ticlopidine inhibitor bound structure (PDB: 3kw4) using YASARA Structure Ver. 14.7.17. The initial binding mode of the respective products (1,3,5-triHX and 1,3,7-triHX) was introduced into the apo-model structures. Based on this initial product bound model, the binding mode was manually built and energy minimized using YASARA (Krieger et al., 2004). To remove bumps and correct the covalent geometry, first a short steepest descent minimization was performed. After removal of conformational stress, the procedure was continued by simulated annealing (timestep 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e. the energy improved by less than 0.05 kJ/mol per atom during 200 steps using the AMBER03 (Duan et al., 2003) force field for protein residues and the general amber force field GAFF (Wang et al., 2004). AM1BCC (Jakalian et al., 2002) calculated partial charges and a force cutoff of 0.786 Å and particle mesh Ewald for exact treatment of long range electrostatics by periodic boundary conditions were employed. The same procedure was applied to the active site mutation S375A/L378S/A483T/M486L/K488R/N489K (mut6).

IV. Results

1. Bioinformatic analysis of *Hypericum calycinum* SSH library

Yeast extract-treated *H. calycinum* cell cultures synthesize a 1,3,7-trihydroxyxanthone derivative known as hyperxanthone E (**Figure IV-1**) which is not detectable in untreated cultures. These conditions were utilized to construct a suppression subtractive hybridization (SSH) cDNA library, where mRNA pools were isolated from elicitor-treated and non-treated cell cultures, converted into cDNA pools and used for subtraction, resulting in a high enrichment of those mRNAs/cDNAs which are elicitor-induced (Gaid et al., 2012).

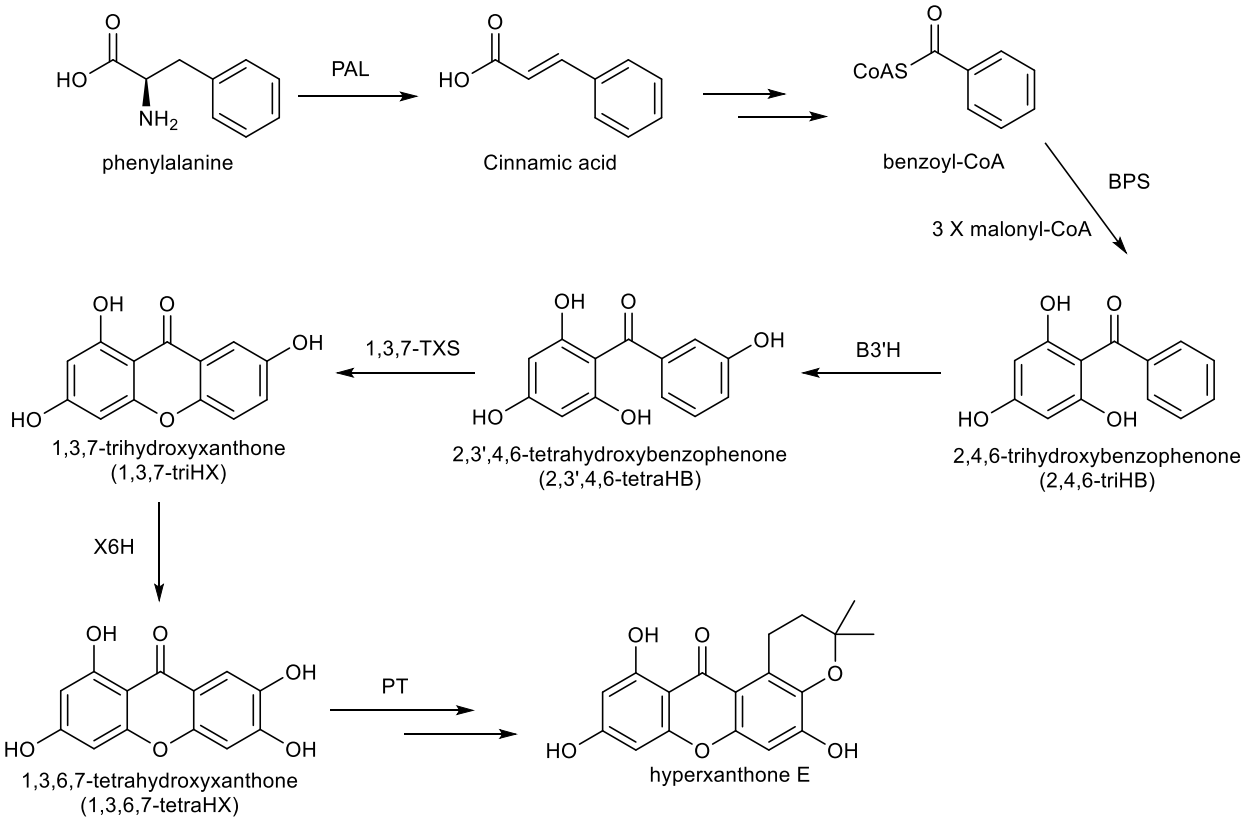


Figure IV-1. Proposed biosynthetic pathway of hyperxanthone E in *H. calycinum* cell cultures. PAL is phenylalanine ammonia lyase, BPS is benzophenone synthase, B3′H is benzophenone 3′ hydroxylase, 1,3,7-TXS is 1,3,7-trihydroxyxanthone synthase, X6H is xanthone 6 hydroxylase, PT is prenyltransferase.

A portion of the SSH library was sequenced and 2005 clones were bioinformatically processed, assembled into 277 contigs using CAP3 program (Huang and Madan, 1999) and functionally annotated using Blast2GO program (Conesa et al., 2005). Contigs annotated to encode CYPs were individually blasted against the non-redundant protein database (NR) of NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). When showing E-values lower than (1.0 E-6) with one or more CYP sequences in the database, contigs were collected and the subfamily of the closest classified CYP as well as the number of copies per contig (retrieved from CAP3 output) were listed in **Table IV-1**. Moreover, the contigs encoding CYPs were blasted against *H. perforatum* MPGR trascriptome database (http://medicinalplantgenomics.msu.edu/mpgr_blast.shtml). For

each contig, the percentage identity with the closest locus was also listed in **Table IV-1**. The assembled contigs represented candidate *CYP* cDNA fragments that encoded proteins belonging to the CYP71, CYP72, CYP81 and CYP706 families. Most transcripts coded for members of the CYP81 and CYP72 families. CYP81 family proteins belong to the CYP71 clan, whose members are known to catalyze various reactions in the shikimate pathway. On the other hand, CYP72 family proteins belong to the CYP72 clan, whose members are mainly recruited for modification of isoprenoids (Nelson et al., 2004).

Table IV-1. *CYP* contigs identified in a *H. calycinum* subtracted cDNA library.

Contig number	Number of copies	Closest NCBI match	Closest MPGR[a] match (% identity)
21	22	CYP81C	hpa_locus_416 (95.0)
41	7	CYP81C	hpa_locus_416 (91.6)
50	11	CYP81C	hpa_locus_416 (97.2)
56	2	CYP71A	hpa_locus_18081 (98.2)
59	33	CYP81C	hpa_locus_416 (97.3)
62	36	CYP72A	hpa_locus_60 (78.2)
86	3	CYP81D	hpa_locus_12223 (70.8)
103	2	CYP81C	hpa_locus_416 (97.4)
110	3	CYP81C	hpa_locus_416 (97.2)
158	2	CYP706	hpa_locus_19031 (81.9)
181	3	CYP706	hpa_locus_19031 (82.2)
199	3	CYP81D	hpa_locus_13843 (89.7)
219	5	CYP81C	hpa_locus_416 (92.0)

[a] Medicinal Plant Genomics Resource (http://medicinalplantgenomics.msu.edu)

2. Cloning of full-length *HcCYP81AA1* cDNA

2.1. Amplification of the core fragment

Among the overexpressed genes in the SSH library, CYP81 candidates were particularly encoded by several contigs, of which contigs 21 and 59 were represented by 22 and 33 copies, respectively (**Table IV-1**). These copy numbers were significantly higher than those for any other CYP that belonged to the shikimate metabolizing CYP71 clan present in the library, indicating the high inducibility upon elicitation. Contigs 21 and 59 shared 95.0 and 97.3% amino acid sequence identity, respectively, with the predicted peptide sequence of hpa_locus_416 from the MPGR *H. perforatum* transcriptome, suggesting that they belong to the same sequence. To test this hypothesis, a forward primer from the upstream contig 59 (Conig59-F1) and a reverse primer from the downstream contig 21 (Contig21-R1) were used and a core fragment of 1075 bp was amplified, confirming that both contigs belonged to the same sequence (**Figure IV-2**). The amplified core fragment lacked stretches of 204 bp toward the start codon and 248 bp toward the stop codon. The missing parts were obtained through Rapid amplification of cDNA ends (RACE), as described subsequently.

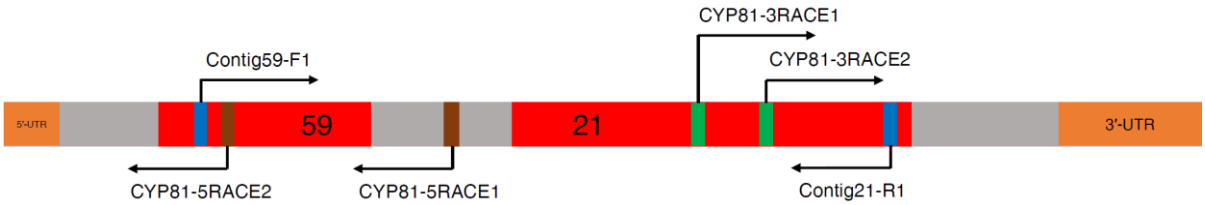


Figure IV-2. Schematic representation of the relative positions of contigs and primers used for the cloning of the full-length cDNA encoding HcCYP81AA1. The contigs obtained from the SSH library are shown in red, the missing parts of the ORF in grey and the UTRs in orange. The primers used to amplify the core fragment are shown in blue, those used for 3′-RACE in green and those employed for 5′-RACE in brown.

2.2. Rapid amplification of cDNA ends (RACE)

Two forward primers (CYP81-3RACE1 and CYP81-3RACE2) were designed based on the sequence of the amplified core fragment. They were used to amplify the remaining 3′ stretch towards the stop codon, in addition to the 3′-UTR and polyA tail by 3′-RACE technique, as previously described (III.2.6.1). Likewise, two reverse primers (CYP81-5RACE1 and CYP81-5RACE2) were designed and the downstream primer (CYP81-5RACE1) was used for transcript-specific reverse transcription. The remaining 5′ stretch together with the 5′-UTR was amplified by 5′-RACE technique, as previously described (III.2.6.3) (**Figure IV-2**).

The complete cDNA comprised 1830 bp, of which 1527 bp constituted the ORF, 89 bp the 5′-UTR, 185 bp the 3′-UTR and 29 bp the polyA tail. The predicted protein product consisted of 508 aa with a predicted molecular mass of 57.1 kDa and a pI of 8.16. The deduced amino acid sequence shared 41 % identity with CYP81K1 from *Arabidopsis thaliana* (NP_196623.1), 45 % with CYP81W1 from *Vitis vinifera* (XP_002285066.1), and 52 % with CYP81C3 from *Populus trichocarpa* (XP_006374861.1), all of which have not been functionally characterized. The protein sequence was submitted to the cytochrome P450 nomenclature committee and was given the name CYP81AA1 (Nelson, 2006). The enzyme constituted a new CYP81 subfamily.

3. Cloning of full-length NADPH-cytochrome P450 reductase *HcCPR* cDNA

The activity of most eukaryotic cytochrome P450 enzymes requires the presence of NADPH-cytochrome P450 reductase (CPR) as a redox partner. CPR receives two electrons from NADPH and transfers them one at a time to the CYP in the course of the catalytic cycle (Jensen and Moller, 2010). Therefore, a full-length CPR transcript from *H. calycinum* cell culture was also cloned.

3.1. Identification of a candidate CPR fragment from the SSH library

In order to identify a core fragment encoding CPR from the SSH library, the TBLASTN algorithm in NCBI (Gertz et al., 2006) was utilized. All ESTs in the library were queried against the amino acid sequence of NADPH-cytochrome P450 reductase 1 from *Arabidopsis thaliana* (AtCPR1) (Mizutani and Ohta, 1998) and the results were ordered according to the calculated maximum score. Only two ESTs that belong to a single contig were identified, they shared about 80 % identity with the amino acid sequence of AtCPR1 (**Figure IV-3**). Therefore, this sequence was

used as a core fragment to clone a full-length CPR transcript from *H. calycinum*. This core fragment lacked 1707 and 244 bp toward the start and stop codons, respectively.

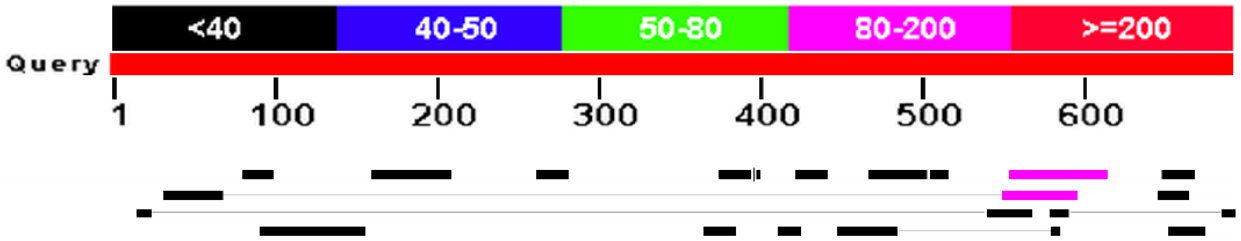


Figure IV-3. Representation of TBLASTN result from the alignment of AtCPR1 protein sequence (NP_194183.1) against the translated nucleotide sequences of all 2005 ESTs from the SSH library, showing in purple the two ESTs that represent the HcCPR2 core fragment. The upper color key represents the alignment score.

3.2. Extension of the CPR core fragment in the 3′ direction

Two forward primers (CPR-3RACE1 and CPR-3RACE2) were designed based on the sequence of the identified core fragment. They were used to amplify by the 3′-RACE technique (III.2.6.1) the remaining 244 bp towards the stop codon, in addition to 309 bp 3′-UTR and 34 bp polyA tail (**Figure IV-4**).

Figure IV-4. Schematic representation of the relative positions of the CPR core fragment and the primers used for the cloning of full-length HcCPR2 cDNA. The core fragment obtained from the SSH library is shown in red, the missing parts of the ORF in grey, the UTRs in orange. The primers used to extend the core fragment towards the 5′-end are shown in blue, those used for 3′-RACE in green and those used for 5′-RACE in brown.

3.3. Extension of the CPR core fragment in the 5′ direction

The identified CPR core fragment lacked 1707 bp towards the start codon. Initial trials to extend it by 5′-RACE (III.2.6.2 or III.2.6.3) were not successful. Therefore, we searched for an upstream conserved region to design a degenerate primer that could be used for partial extension of the fragment into the 5′ direction. Multiple alignment of the nucleotide sequences of 20 plant CPRs collected from the NCBI database was created using ClustalW (Thompson et al., 1994). An upstream forward degenerate primer for CPRs was designed based on the sequence of the conserved FMN binding domain observed in the alignment (**Figure IV-5**). Standard PCR utilizing the designed degenerate primer (CPR-dpF-1) with a reverse gene-specific primer (CPR-R1) and 5′ RACE-ready cDNA as a template led to extension of the core fragment 1362 bp in the 5′ direction (**Figure IV-4**).

Subsequently, two reverse primers (CPR-5RACE1 and CPR-5RACE2) were designed based on the sequence of the extended fragment. The CPR-5RACE1 primer was used for CPR transcript-specific reverse transcription and the remaining 5′ stretch together with the 5′-UTR was amplified

by 5′-RACE (III.2.6.3). The complete cDNA comprised 2599 bp, of which 2139 bp constituted the ORF, 117 bp the 5′-UTR, 309 bp the 3′-UTR and 34 bp the polyA tail. The predicted protein product consisted of 712 aa with a predicted molecular mass of 78.7 kDa and a pI of 5.39. The deduced amino acid sequence shared 75, 79.1 and 79.2% identity with the CPR2 enzymes from *Arabidopsis thaliana* (Mizutani and Ohta, 1998), *Gossypium hirsutum* (Yang et al., 2010), and *Populus trichocarpa* (Ro et al., 2002), respectively.

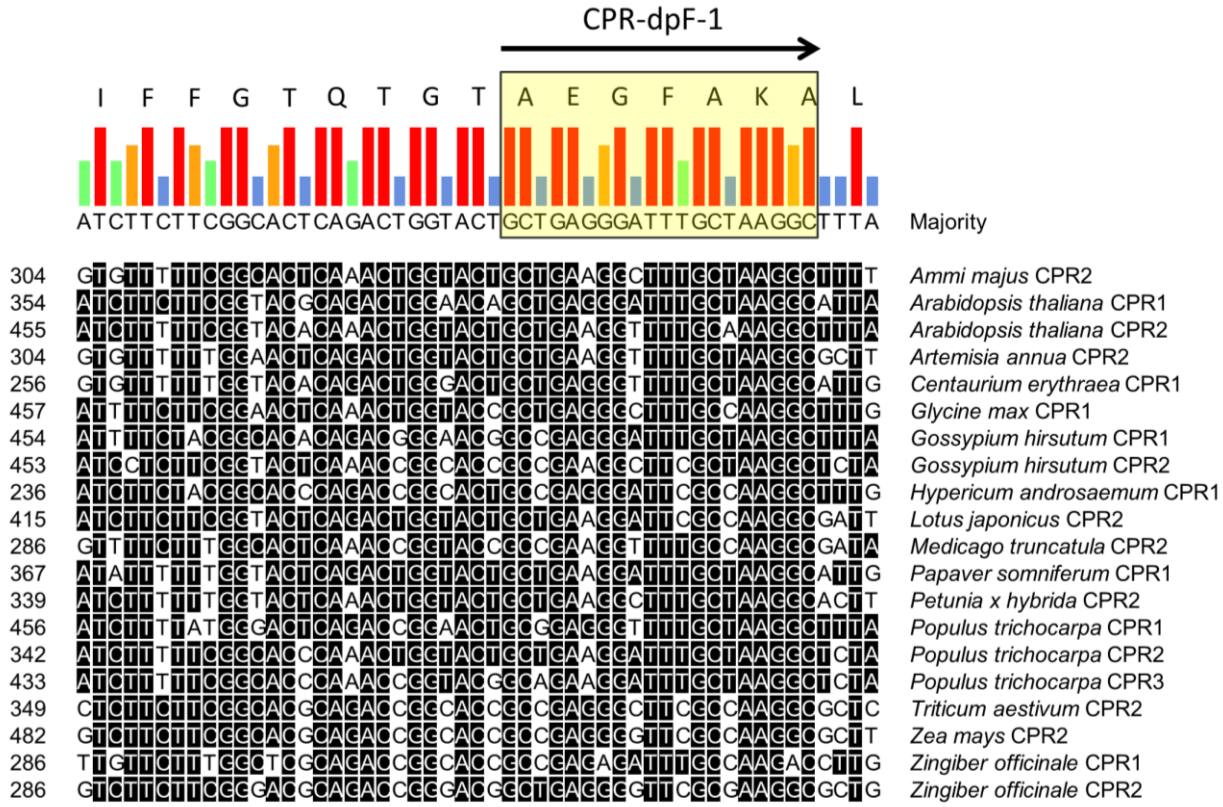


Figure IV-5. Multiple sequence alignment of the nucleotide sequences corresponding to FMN binding domains of 20 plant CPRs for designing the upstream degenerate primer (CPR-dpF-1), which was used to extend the CPR core fragment in the 5′ direction. The colored bars represent consensus strength. *Ammi majus* CPR2 (AY532374.1), *Arabidopsis thaliana* CPR1 (NM_118585.3), *A. thaliana* CPR2 (NM_179141.2), *Artemisia annua* CPR2 (DQ984181.1), *Centaurium erythraea* CPR1 (AY596976.1), *Glycine max* CPR1 (NM_001249813.1), *Gossypium hirsutum* CPR1 (FJ719368.1), *G. hirsutum* CPR2 (FJ719369.1), *Hypericum androsaemum* CPR1 (AY520902.1), *Lotus japonicus* CPR2 (AB433810.1), *Medicago truncatula* CPR2 (XM_003610061.1), *Papaver somniferum* CPR1 (U67185.1), *Petunia* x *hybrida* CPR2 (DQ099545.1), *Populus trichocarpa* CPR1 (XM_002307300.2), *P. trichocarpa* CPR2 (AF302497.1), *P. trichocarpa* CPR3 (AF302498.1), *Triticum aestivum* CPR2 (AJ303373.1), *Zea mays* CPR2 (AJ303373.1), *Zingiber officinale* CPR1 (AB566408.1), *Z. officinale* CPR2 (AB566409.1).

4. Analysis of the HcCPR2 amino acid sequence

The deduced amino acid sequence of HcCPR2 was aligned with the CPR homologs from *Arabidopsis thaliana* (AtCPR1 and 2) and *Gossypium hirsutum* (GhCPR1 and 2) (**Figure IV-6**). All conserved functional domains characteristic for CPRs, which included FMN-, FAD-, NADPH- and P450-binding domains, were identified in HcCPR2 (**Figure IV-6**). An N-terminal hydrophobic domain spanning the region between amino acids 50-70 was also identified. It acts

as a membrane anchor attaching HcCPR2 to the endoplasmic reticulum, as previously proposed for other CPRs (Mizutani and Ohta, 1998; Yang et al., 2010).

Figure IV-6. Multiple sequence alignment of the deduced amino acid sequence of HcCPR2 with *Arabidopsis* and *Gossypium* CPR homologs showing the common conserved domains characteristic for CPRs. Conserved amino acid residues in all five CPRs are shown in white on a red background.

5. Functional characterization of HcCPR2 in yeast cells

5.1. Construction of the expression plasmid and protein expression

For functional expression, the HcCPR2 ORF was amplified with a proofreading polymerase using yeast-elicited *H. calycinum* cDNA as a template (**Figure IV-7a**) and cloned between *Bam*HI and *Hin*dIII restriction sites in multiple cloning site 2 (MCS2) of the pESC-URA vector (**Figure IV-7b**), as described under (III.2.3). The constructed plasmid was transformed into INV*Sc*1 yeast cells (III.3.3.3) and transcription was commenced under the effect of GAL1 promoter as described under (III.4.1). The crude microsomal fraction was isolated (III.4.2) and used for enzymatic incubations.

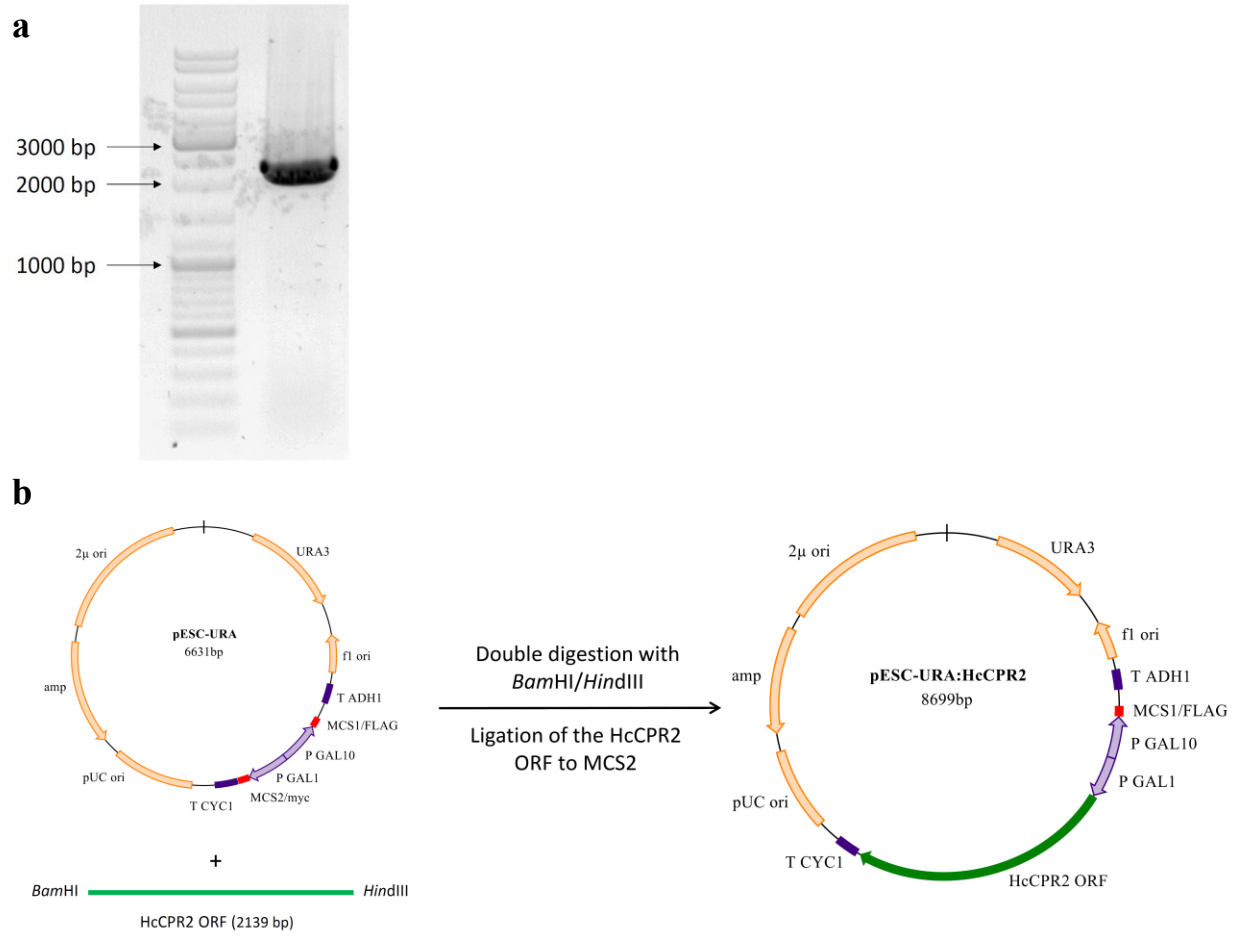


Figure IV-7. (**a**) Amplification of full-length HcCPR2 with a proofreading polymerase. (**b**) Construction of the pESC-URA:HcCPR2 plasmid by cloning the HcCPR2 ORF in MCS2 of the pESC-URA yeast expression vector.

5.2. Reduction of cytochrome c (Cyt c)

The activity of HcCPR2 was examined through its ability to transfer electrons from NADPH to Cyt c as described in III.4.5.1. HcCPR2 reduced cytochrome c, as shown by the increase of absorbance at 550 nm directly after the addition of NADPH (**Figure IV-8**). By measuring the reaction rates at various Cyt c concentrations ranging from 1–300 µM, the Km value for HcCPR2

with cytochrome c was determined to be 35.1 ± 5.7 µM (**Figure IV-9a**). Similarly, the reaction rates were measured at various NADPH concentrations ranging from 5–500 µM and the Km value was determined to be 43.9 ± 6.5 µM (**Figure IV-9b**). Therefore, this enzyme could be used as a redox partner for the cloned CYP.

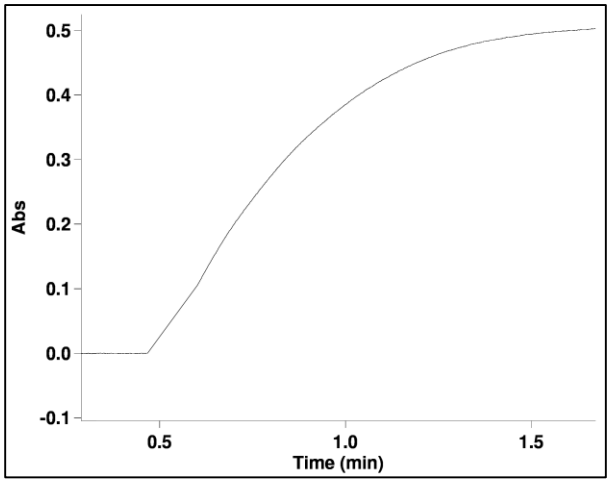


Figure IV-8. Cyt c reduction assay showing increased absorbance at 550 nm upon addition of NADPH.

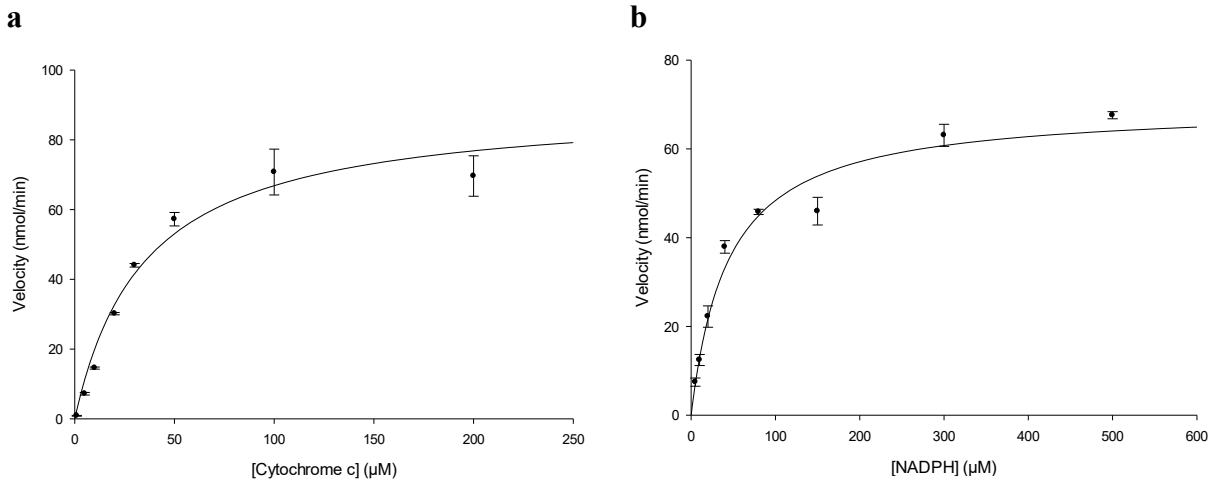


Figure IV-9. Michaelis-Menten hyperbolic regression curves for the determination of Km values of HcCPR2 with (**a**) cytochrome c and (**b**) NADPH.

6. Functional characterization of HcCYP81AA1

6.1. Construction of the expression plasmid and protein expression

The HcCYP81AA1 ORF was amplified with a proofreading polymerase from the cDNA obtained from yeast-treated *H. calycinum* cell culture (**Figure IV-10a**) and cloned between *Eco*RI and *Pac*I restriction sites in MCS1 of the pESC-URA:HcCPR2 plasmid (**Figure IV-10b**), as described in III.2.3. The constructed expression plasmid containing both ORFs encoding HcCYP81AA1 and

HcCPR2 was transferred into INV*Sc*1 yeast cells as described in III.3.3.3. Co-expression of HcCYP81AA1 and HcCPR2 was induced under the effect of GAL10 and GAL1 promoters, respectively. The microsomal fraction was isolated (III.4.2) and used for enzymatic incubations.

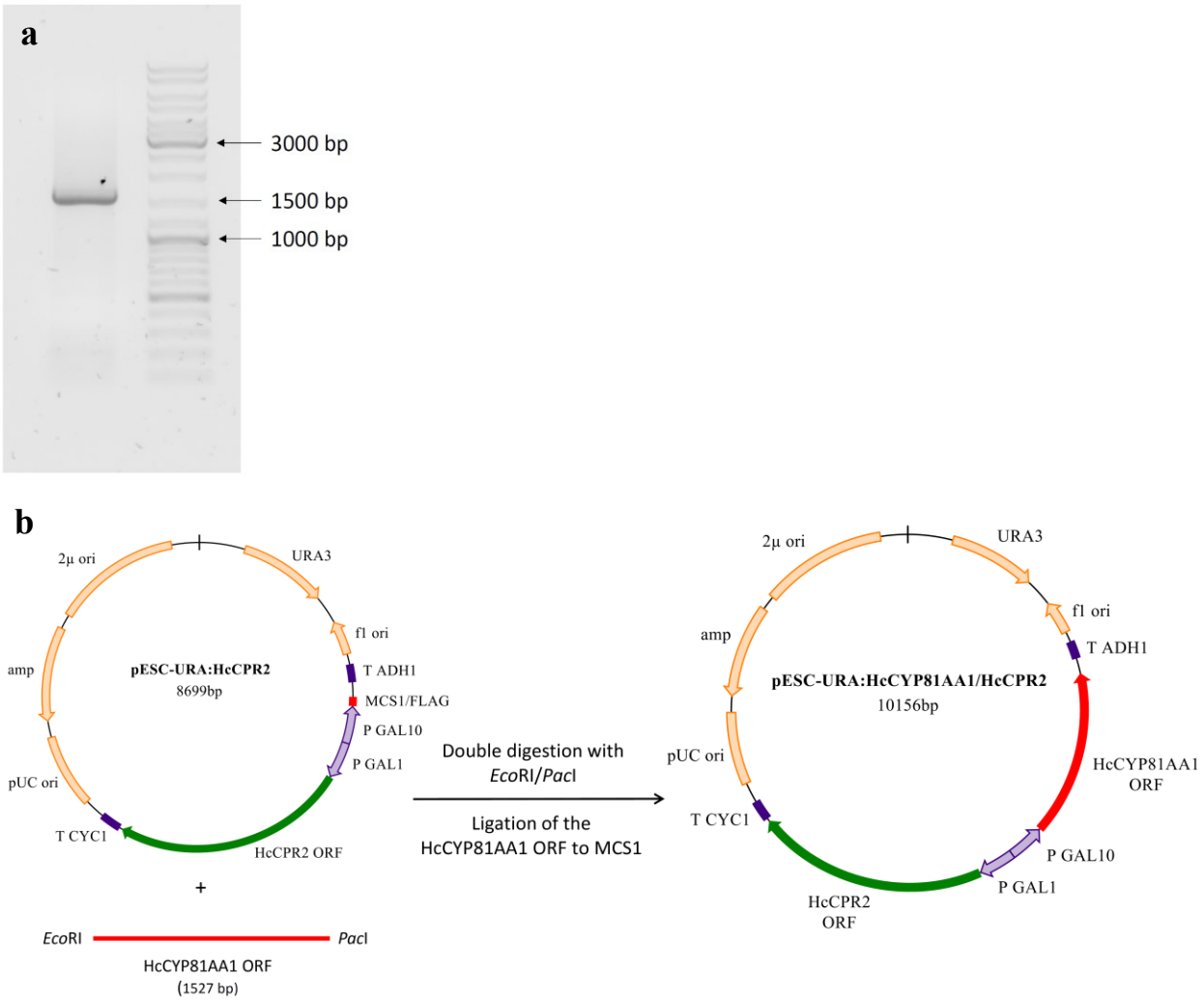


Figure IV-10. (**a**) Amplification of full-length HcCYP81AA1 with a proofreading polymerase. (**b**) Construction of the pESC-URA:HcCYP81AA1/HcCPR2 plasmid by cloning the HcCYP81AA1 ORF in MCS1 of the pESC-URA:HcCPR2 plasmid.

6.2. Enzymatic activities of HcCYP81AA1

The isolated yeast microsomes expressing both HcCYP81AA1 and HcCPR2 were incubated with eleven potential benzophenone and xanthone substrates (**Figure IV-11**) in the presence of NADPH (III.4.5.2). The products formed were analyzed by HPLC-DAD and LC-MS (III.5.1, III.5.2).

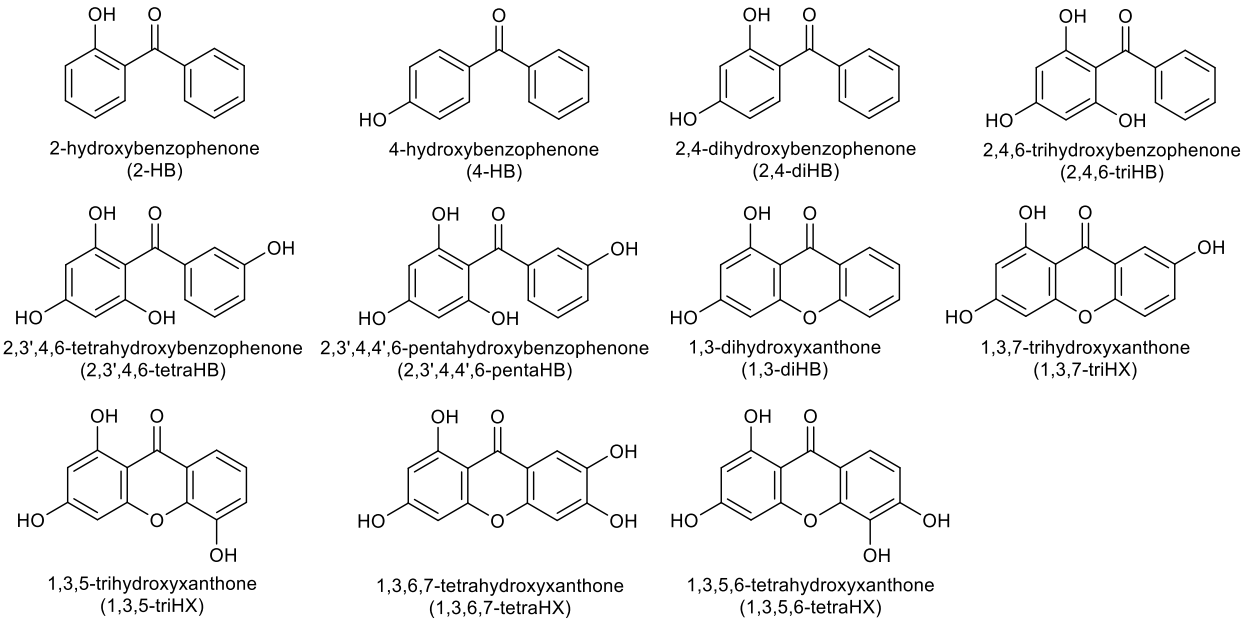


Figure IV-11. Potential benzophenone and xanthone substrates.

Incubation with 2,3′,4,6-tetrahydroxybenzophenone (2,3′,4,6-tetraHB) resulted in the formation of a single product (R_t 19.6 min) (**Figure IV-12a**). This product was identified as 1,3,7-trihydroxyxanthone (1,3,7-triHX). Its UV spectrum showed four characteristic absorption maxima at 234, 257, 309 and 374 nm, which matched those of authentic 1,3,7-triHX (**Figure IV-12b**). In addition, its MS spectrum showed a molecular ion peak at m/z 245 equivalent to $[M + H]^+$ of trihydroxylated xanthones (**Figure IV-12c**) and the fragmentation pattern agreed with that of the reference compound.

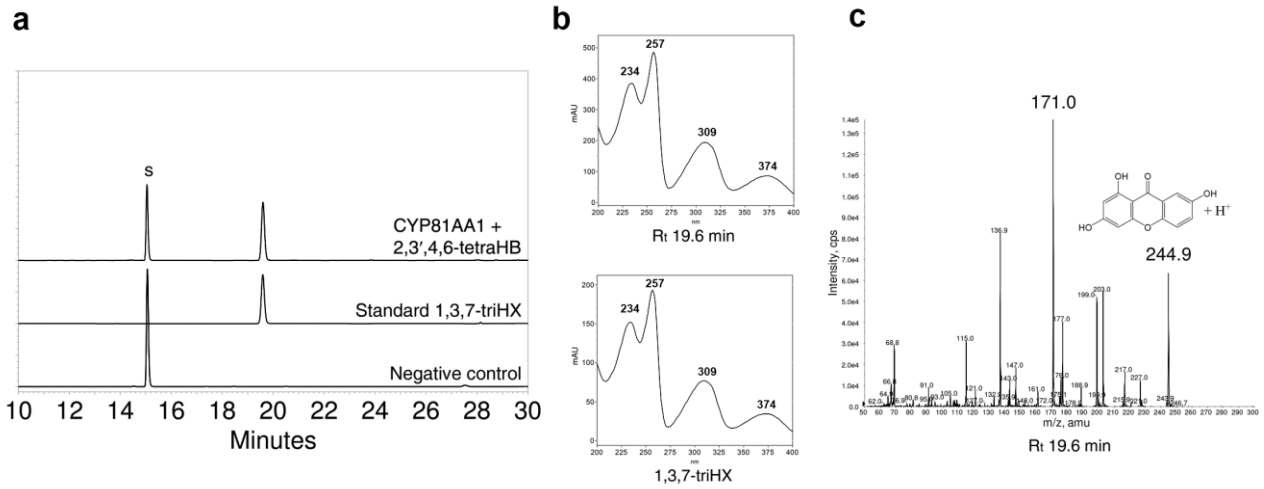


Figure IV-12. Incubation of the microsomes containing HcCYP81AA1 and HcCPR2 with 2,3′,4,6-tetraHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standard, and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic product and standard. (**c**) EPI mass spectrum of the enzymatic product.

Incubation with 2,4,6-trihydroxybenzophenone (2,4,6-triHB) led to the formation of two products with R_t values of 15.1 and 19.6 min, respectively (**Figure IV-13a**). The product with R_t 19.6 min was identified as 1,3,7-triHX, as described in the previous incubation. The product with R_t 15.1 min was identified as 2,3′,4,6-tetraHB by its UV and MS spectra. The UV spectrum matched that of authentic 2,3′,4,6-tetraHB and showed absorption maxima at 205, 258 and 305 nm (**Figure IV-13b**). EPI-MS revealed a molecular ion peak at m/z 247 equivalent to $[M + H]^+$ of tetrahydroxylated benzophenones. Beside the molecular ion peak, two other prominent peaks were identified (m/z 121 and 153). They correspond to the ions resulting from breaking the bond at either side of the carbonyl group, thereby confirming the presence of three hydroxyl groups on the first aromatic ring (m/z 153) and one hydroxyl group on the second ring (m/z 121) of the benzophenone skeleton (**Figure IV-13c**).

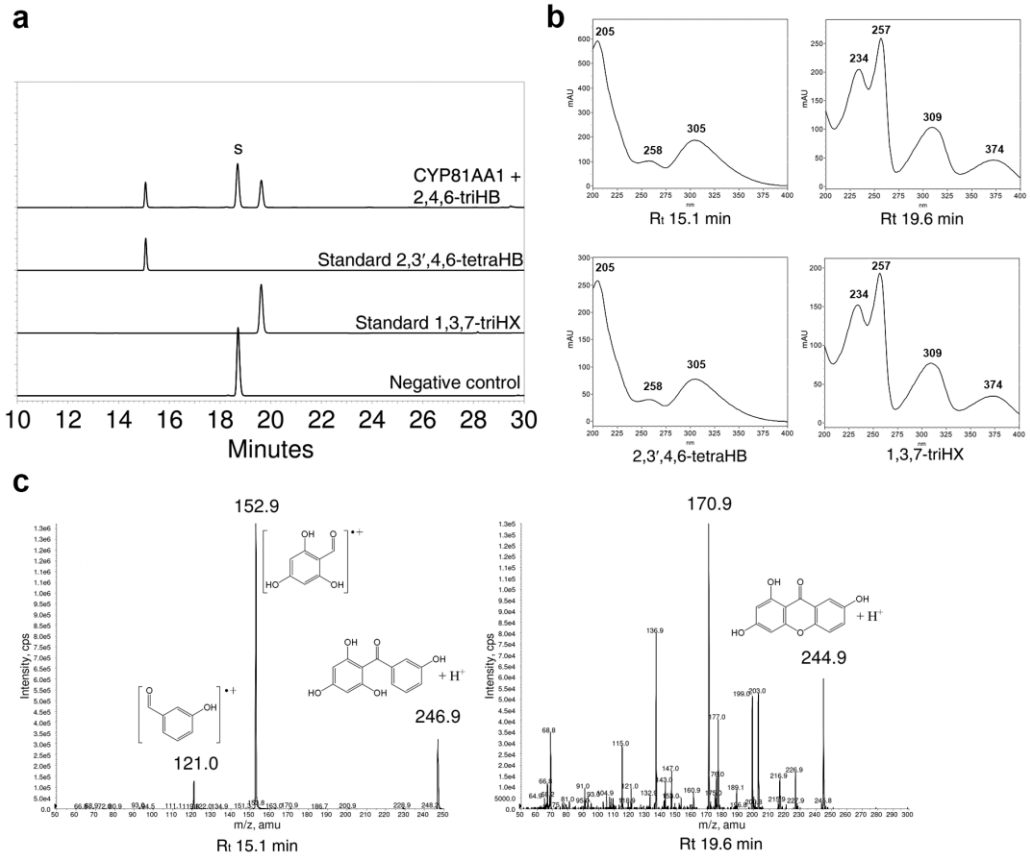


Figure IV-13. Incubation of the microsomes containing HcCYP81AA1 and HcCPR2 with 2,4,6-triHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standards and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic products and standards. (**c**) EPI mass spectra of the enzymatic products.

Similarly, incubation with 2,4-dihydroxybenzophenone (2,4-diHB) yielded two products with R_t values of 13.7 and 17.1 min, respectively (**Figure IV-14a**). The product with R_t 13.7 min was identified as 2,2′,4,5′-tetrahydroxybenzophenone (2,2′,4,5′-tetraHB) by its UV and MS spectra. The UV spectrum matched that of authentic 2,2′,4,5′-tetraHB and showed absorption maxima at 215 and 282 nm (**Figure IV-14b**). EPI-MS revealed a molecular ion peak at m/z 247 equivalent to $[M + H]^+$ of tetrahydroxylated benzophenones. Beside the molecular ion peak, only one prominent peak was identified in the mass spectrum (m/z 137). It corresponds to the ions resulting from breaking the bond at either side of the carbonyl group, thereby confirming the presence of two hydroxyl groups on each aromatic ring of the benzophenone skeleton (**Figure IV-14c**).The R_t 17.1 product was tentatively identified by MS as 2,3′,4-trihydroxybenzophenone (2,3′,4-triHB), as a standard was not available. EPI-MS revealed a molecular ion peak at m/z 231 equivalent to $[M + H]^+$ of trihydroxylated benzophenones. The fragmentation pattern confirms the presence of two hydroxyl groups on the first aromatic ring (m/z 137) and one hydroxyl group on the second ring (m/z 121) of the benzophenone skeleton (**Figure IV-14c**).

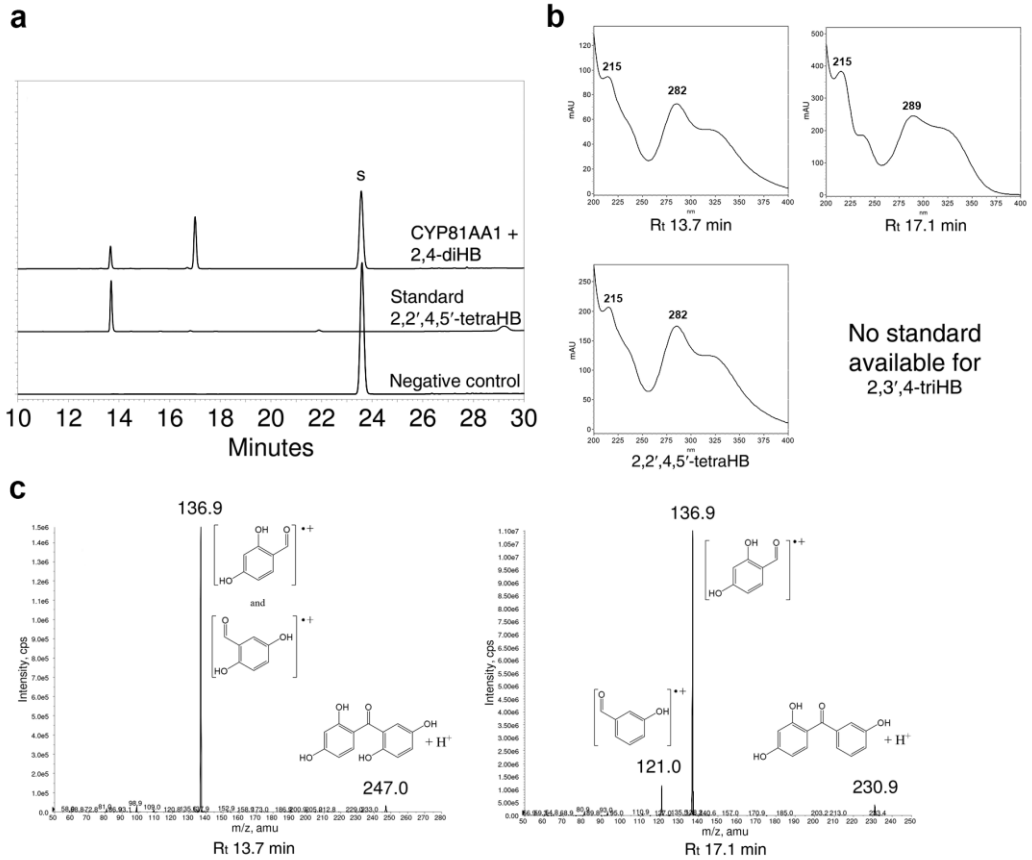


Figure IV-14. Incubation of the microsomes containing HcCYP81AA1 and HcCPR2 with 2,4-diHB. (**a**) Stacked HPLC-DAD chromatograms for the incubation, the available standard and the negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic products and available standard. (**c**) EPI mass spectra of the enzymatic products.

Incubation with 2,3′,4,4′,6-pentahydroxybenzophenone (2,3′,4,4′,6-pentaHB, maclurin) yielded a single product (R_t 22.4 min) (**Figure IV-15a**). This compound was identified as 1,3,6,7-tetrahydroxyxanthone (1,3,6,7-tetraHX, norathyriol) by its UV spectrum, which showed four characteristic absorption maxima at 236, 253, 314 and 363 nm matching those of authentic 1,3,6,7-tetraHX (**Figure IV-15b**). In addition, the MS spectrum showed a molecular ion peak at m/z 261 equivalent to $[M + H]^+$ of tetrahydroxylated xanthones (**Figure IV-15c**), the fragmentation pattern agreeing with that of the reference compound. Traces of 1,3,6,7-tetraHX were also detected in the negative control. The presence of 1,3,6,7-tetraHX as a natural contaminant in 2,3′,4,4′,6-pentaHB samples was previously described (Jefferson and Scheinmann, 1965).

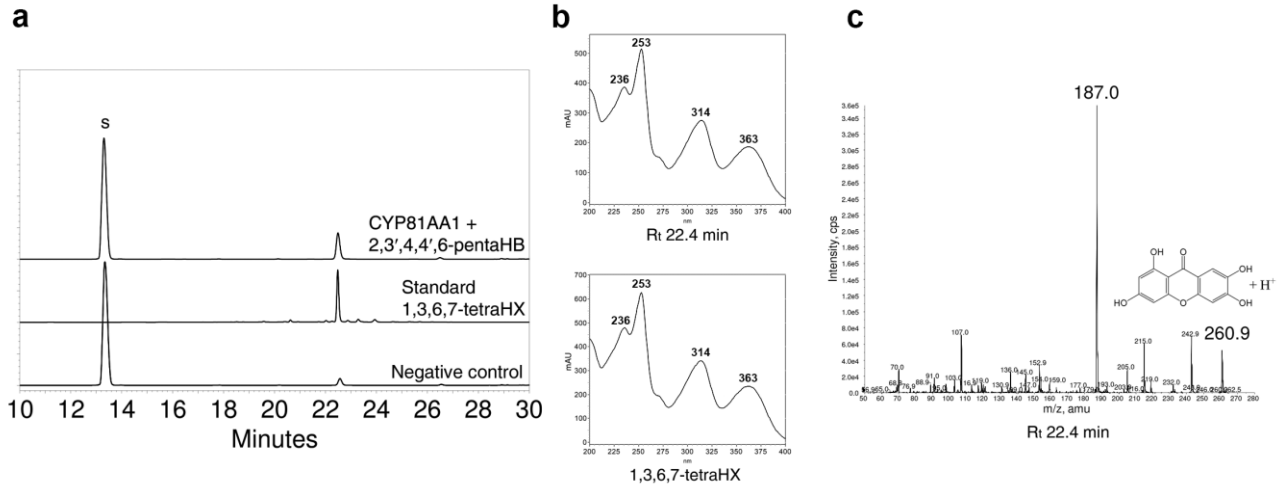
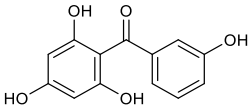
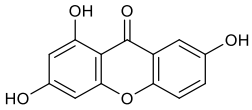
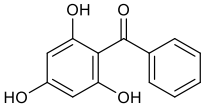
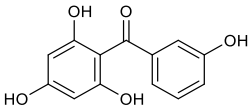
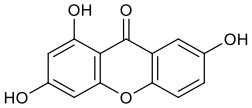
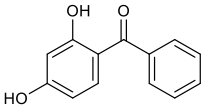
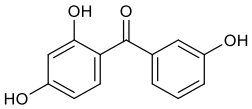
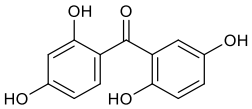
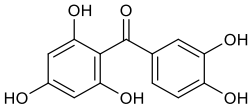
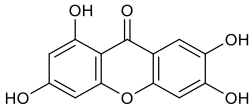


Figure IV-15. Incubation of the microsomes containing HcCYP81AA1 and HcCPR2 with 2,3′,4,4′,6-pentaHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standard and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic product and standard. (**c**) EPI mass spectrum of the enzymatic product.

For all substrates, no enzymatic products were detectable in assays lacking NADPH. Likewise, microsomes from yeast cells harboring the empty plasmid failed to produce any of the products. No activity was detectable with 2-hydroxybenzophenone (2-HB), 4-hydroxybenzophenone (4-HB), 1,3-dihydroxyxanthone (1,3-diHX), 1,3,7-trihydroxyxanthone (1,3,7-triHX), 1,3,5-trihydroxyxanthone (1,3,5-triHX), 1,3,5,6-tetrahydroxyxanthone (1,3,5,6-tetraHX) and 1,3,6,7-tetraHX.

In conclusion, activity was detected with four benzophenones of the eleven benzophenone and xanthone substrates tested (**Table IV-2**). Accordingly, HcCYP81AA1 was identified as a bifunctional enzyme possessing both benzophenone 3′-hydroxylase (B3′H) and 1,3,7-trihydroxyxanthone synthase (1,3,7-TXS) activities.

Table IV-2. Summary of the enzymatic activities of HcCYP81AA1.

Substrate	Product(s)
 2,3′,4,6-tetraHB	 1,3,7-triHX
 2,4,6-triHB	 2,3′,4,6-tetraHB  1,3,7-triHX
 2,4-diHB	 2,3′,4-triHB  2,2′,4,5′-tetraHB
 2,3′,4,4′,6-pentaHB	 1,3,6,7-tetraHX
2-HB	No product detected
4-HB	No product detected
1,3-diHX	No product detected
1,3,7-triHX	No product detected
1,3,5-triHX	No product detected
1,3,6,7-tetraHX	No product detected
1,3,5,6-tetraHX	No product detected

6.3. Improvement of incubation conditions

6.3.1. Effect of pH and temperature

The pH value of the surrounding solution affects the protein folding through changing the degree of ionization of its acidic and basic amino acid residues. Therefore, enzymes are usually active within a limited range of pH values with highest activity achieved at the optimum pH. Likewise, the enzyme activity increases with increasing incubation temperature until a certain limit where the hydrogen bonds start to break, leading to disruption of the tertiary and quaternary structures; the enzyme starts to denature.

2,3′,4,6-TetraHB was used as a substrate to determine the optimal pH and temperature for HcCYP81AA1. The optimum pH was 7.0 with loss of 46.7 and 30.7 % activity at pH 6.5 and 7.5, respectively (**Figure IV-16a**). The optimum temperature was 30 °C with loss of 18.3 and 24.3 % activity at 25 and 35 °C, respectively (**Figure IV-16b**).

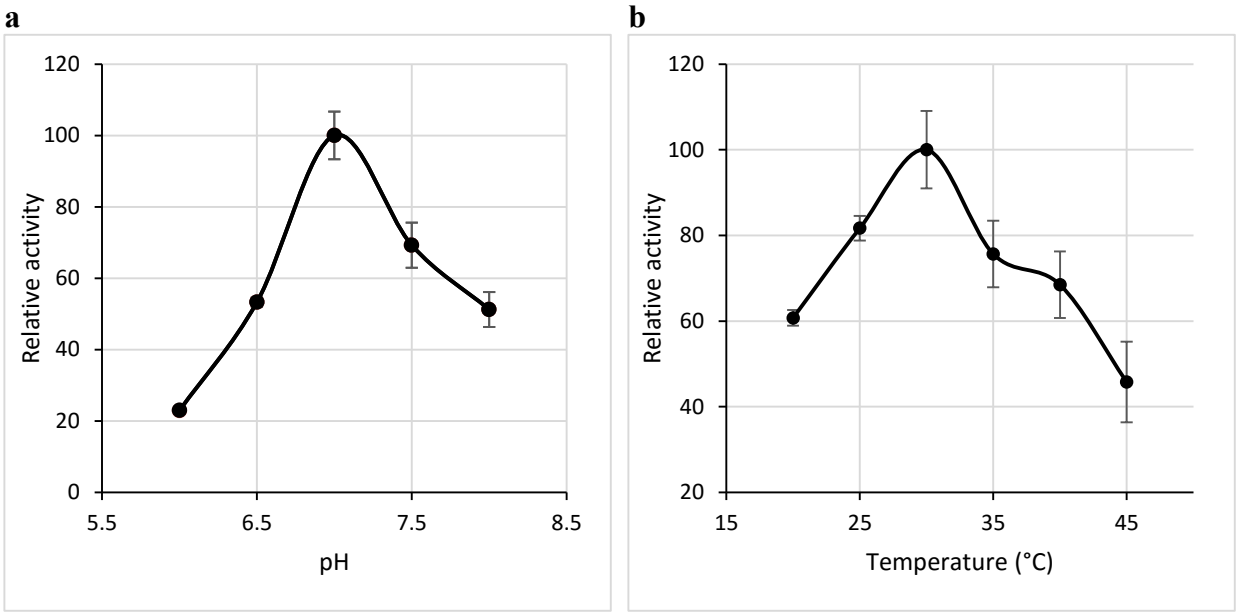


Figure IV-16. Effect of pH and temperature on the activity of HcCYP81AA1. Data represent means ± s.d. (n = 3).

6.3.2. Effect of incubation time and protein amount

The accumulation of 1,3,7-triHX upon incubation with 2,3′,4,6-tetraHB increased over the period of 120 min tested. However, the product formation rate was linear only within the first 15 min of incubation (**Figure IV-17a**). With increasing protein amounts, product accumulation was linear over the whole range tested, i.e. between 6–96 µg microsomal proteins in the standard assay (**Figure IV-17b**).

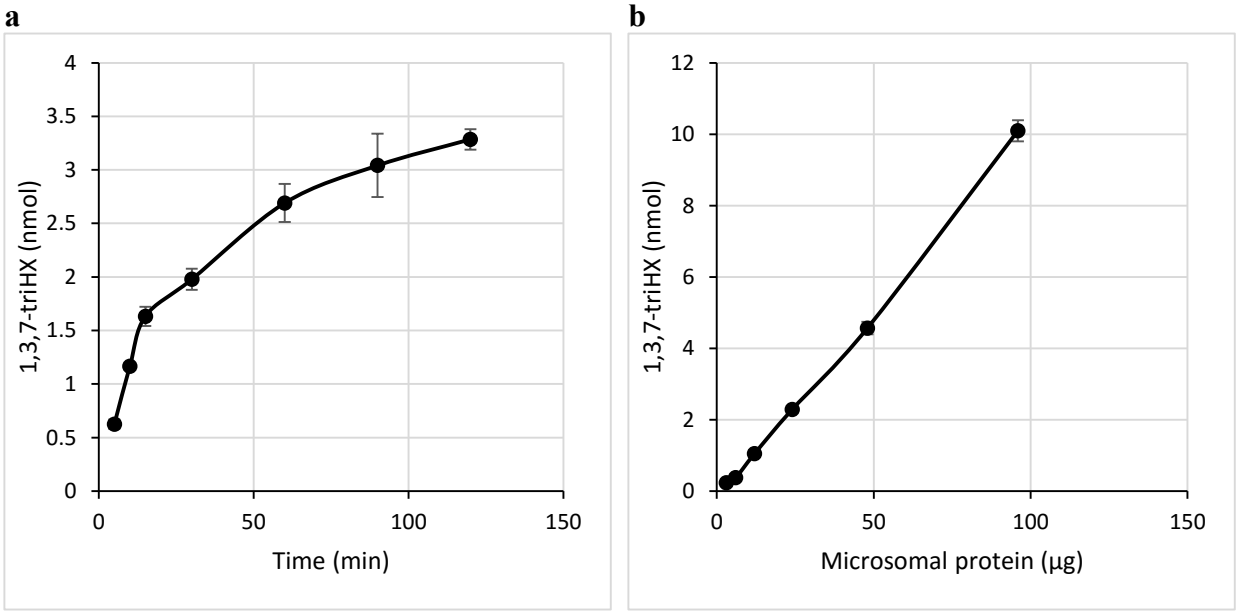


Figure IV-17. Effect of incubation time and protein amount on the activity of HcCYP81AA1. Data represent means ± s.d. (n = 3).

7. Elicitor-induced expression of *HcCYP81AA1* in *H. calycinum* cell cultures

The expression profiles of *HcBPS*, *HcCPR2* and *HcCYP81AA1* were analyzed by quantitative real-time RT-PCR. RNA was isolated from untreated *H. calycinum* cell cultures (0 h) as well as 4, 8, 12, 16, 20, 24, 36 and 48 h after elicitation with 3 g/l of yeast extract. The purity of the isolated RNA was checked by measuring the A260/A280 ratio. All samples showed a ratio of ~2.1 indicating pure RNA. Moreover, the integrity of the isolated RNA samples was verified using agarose gel electrophoresis. Two distinct bands corresponding to 28S and 18S rRNA were observed with an approximate 28S:18S ratio of 2:1 (**Figure IV-18**).

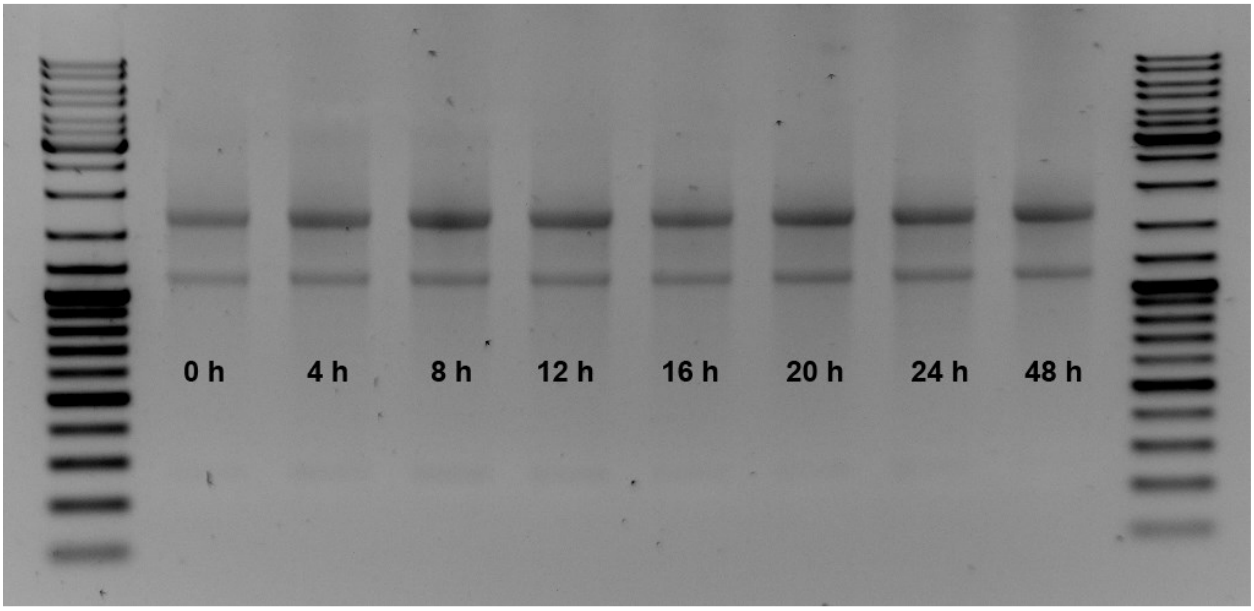


Figure IV-18. Agarose gel-electrophoresis of 3 µg each of the RNA samples used for real-time qRT-PCR.

Primer pairs were designed to amplify fragments of 75–150 bp from the genes under investigation as well as the reference genes using Beacon Designer™ software (PREMIER biosoft, California, USA) and were rechecked with Primer 3.0 software (http://primer3.ut.ee/). Before carrying out the expression analysis experiment, the amplification efficiencies of the designed primers were first determined from the slope of the standard curve based on the Cq values of a serial dilution of the pooled cDNA from each time point. The equation $E=10^{(-1/slope)}$ was used to calculate the efficiencies, while the efficiency in percentage was equal to (E - 1) x 100 (Pfaffl, 2001). All designed primers had an efficiency in the range of 90–110 % (**Table IV-3**).

Table IV-3. Amplification efficiency of qPCR with the targets studied.

Target name	Efficiency percentage	R^2 value	Standard curve slope
HcCYP81AA1	91.5 %	0.998	-3.544
HcBPS	93.2 %	0.999	-3.497
HcCPR2	99.0 %	0.997	-3.346
Actin	100.3 %	0.999	-3.315
Histone H2A	100.9 %	0.998	-3.299

Furthermore, the specificity of each primer pair was checked from the melt curve and by running the final amplification products on a 2 % agarose gel, where a single band was observed in every reaction. No amplification was found with the no-template controls (NTC) and the no-reverse transcription controls (NRT). *Actin* and *Histone H2A* served as reference genes

The qRT-PCR results show that all the tested genes had similar expression profiles. Their transcripts levels increased rapidly reaching a maximum at 8 h post-elicitation and decreased afterwards. The increase in *HcBPS* and *HcCPR2* transcription levels did not exceed 8-fold, while expression of *HcCYP81AA1* was more strongly induced reaching about 16-fold (**Figure IV-19**). These expression profiles come in accordance with the previously published data for hyperxanthone E accumulation in yeast-treated *H. calycinum* cell cultures (Gaid et al., 2012), where the product accumulated in the range between 12–24 h after elicitation. The phenylalanine ammonia-lysae (PAL) and cinnamate:CoA ligase (CNL) genes responsible for upstream steps in xanthone biosynthesis followed the same expression profile (Gaid et al., 2012).

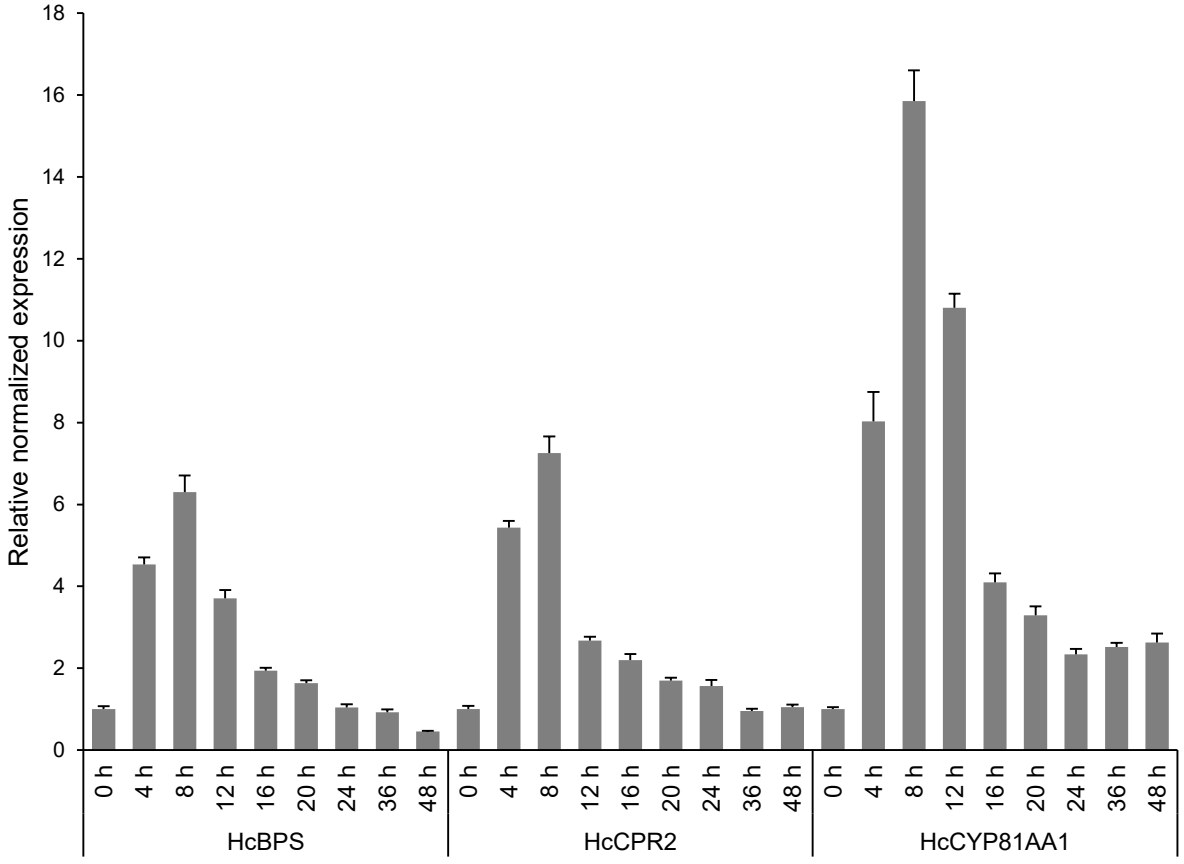


Figure IV-19. Changes in the transcript levels of *HcBPS*, *HcCPR2* and *HcCYP81AA1* in elicitor-treated *H. calycinum* cell cultures relative to those in untreated control cultures. Relative normalized expression was determined by qRT-PCR using *Actin* and *Histone H2A* as reference genes. Data represent means ± s.e.m (n = 3).

8. Identification of HcCYP81AA1 homologs in transcriptomes of *H. perforatum*

H. perforatum was reported to synthesize both 1,3,7- and 1,3,5-triHX derivatives (Crockett et al., 2011; Tocci et al., 2011), indicating occurrence of *1,3,7-TXS* and *1,3,5-TXS* genes. The HcCYP81AA1 sequence was used to search for homologs in the MPGR databank (http://medicinalplantgenomics.msu.edu/). Four closely related CYP loci (416, 928, 8128 and 51544) were identified, which shared 95.3, 70.6, 66.4 and 67.0% amino acid sequence identity with HcCYP81AA1, respectively. The locus 51544 lacked about 430 bp toward the stop codon. All trials to complete the sequence failed, suggesting a pseudogene. Notably, expression of this gene was low in all organs studied, as indicated by the ′fragments per kilobase per million mapped fragments′ (FPKM) values, which are available in the MPGR database. In contrast, the other three loci existed as full-length sequences in the database. The FPKM values of the loci 416 and 928 indicated preferential expression in roots and leaves, which resembles the tissue distribution pattern of BPS transcripts (locus 13339) (**Figure IV-20**). In contrast, transcripts encoded by locus 8128 were more abundant in flower organs.

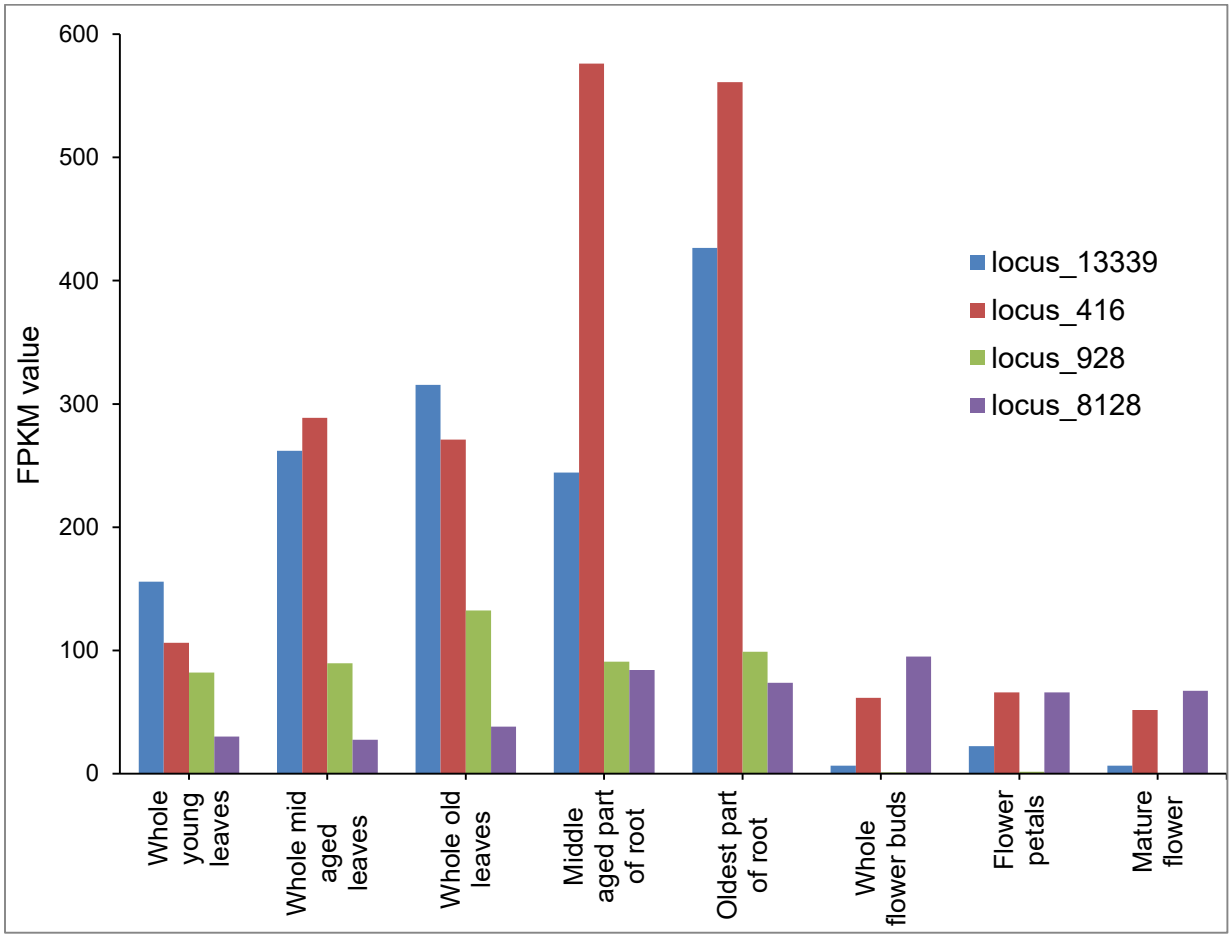


Figure IV-20. Tissue distribution of transcripts encoded by the loci 13339 (*HpBPS*), 416 (*HpCYP81AA1*), 928 (*HpCYP81AA2*) and 8128 (*HpCYP81AA3*). Data are taken from the MPGR database (http://medicinalplantgenomics.msu.edu). FPKM, fragments per kilobase per million mapped fragments.

9. Functional characterization of locus 416 from *H. perforatum* transcriptome

9.1. Construction of the expression plasmid and protein expression

Primers were designed based on the sequence of 'hpa_locus_416_iso_3_len_1885_ver_2' from the MPGR databank. The cDNA obtained from *H. perforatum* mid-aged leaves was used as a template to amplify the ORF of locus 416 with a proofreading polymerase (**Figure IV-21**). It was subsequently cloned between *Spe*I and *Pac*I restriction sites in MCS1 of the pESC-URA:HcCPR2 plasmid as described in III.2.3. The deduced amino acid sequence of locus 416 was referred to as HpCYP81AA1, an ortholog of *H. calycinum* 1,3,7-TXS. The two proteins differed in only 24 amino acids. The constructed expression plasmid containing both ORFs encoding HpCYP81AA1 and HcCPR2 was transferred into INV*Sc*1 yeast cells as described in III.3.3.3. After induction of co-expression, the microsomal fraction was isolated (III.4.2) and used for enzymatic incubations.

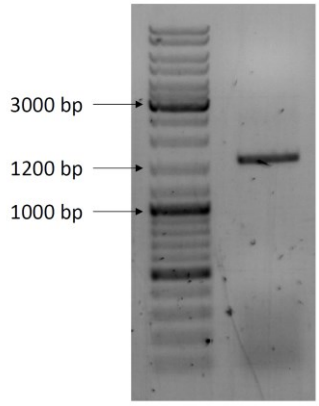


Figure IV-21. Amplification of the full-length HpCYP81AA1 cds (locus 416) from *H. perforatu*m cDNA using a proofreading polymerase.

9.2. Enzymatic activities of HpCYP81AA1 (locus 416)

The isolated yeast microsomes containing both HpCYP81AA1 and HcCPR2 were incubated with the above-mentioned list of potential benzophenone and xanthone substrates (**Figure IV-11**) in the presence of NADPH and the products formed were analyzed by HPLC-DAD and LC-MS. The activity profile of this enzyme resembled that of HcCYP81AA1 as a bifunctional enzyme possessing both B3′H and 1,3,7-TXS activities (**Table IV-4**).

Table IV-4. Summary of the enzymatic activities of HpCYP81AA1.

Substrate	Product(s)	
2,3′,4,6-tetraHB	1,3,7-triHX	
2,4,6-triHB	2,3′,4,6-tetraHB	1,3,7-triHX
2,4-diHB	2,3′,4-triHB	2,2′,4,5′-tetraHB
2,3′,4,4′,6-pentaHB	1,3,6,7-tetraHX	
2-HB	No product detected	
4-HB	No product detected	
1,3-diHX	No product detected	
1,3,7-triHX	No product detected	
1,3,5-triHX	No product detected	
1,3,6,7-tetraHX	No product detected	
1,3,5,6-tetraHX	No product detected	

10. Functional characterization of locus 928 from *H. perforatum* transcriptome

10.1. Construction of the expression plasmid and protein expression

Similar to the previous locus, primers were designed according to the sequence 'hpa_locus_928_iso_9_len_1841_ver_2' from the MPGR databank. The cDNA obtained from *H. perforatum* mid-aged leaves was used as a template to amplify the ORF of locus 928 with a proofreading polymerase (**Figure IV-22**). It was subsequently cloned between *Eco*RI and *Pac*I restriction sites in MCS1 of the pESC-URA:HcCPR2 plasmid (III.2.3). The enzyme encoded by the ORF of locus 928 was designated by the Cytochrome P450 Nomenclature Committee (Nelson, 2006) as HpCYP81AA2, a paralog of CYP81AA1. The constructed expression plasmid (pESC-URA:HpCYP81AA2/HcCPR2) was transferred into INV*Sc*1 yeast cells (III.3.3.3). Following induction of co-expression, the microsomal fraction was isolated (III.4.2) and used for enzymatic incubations.

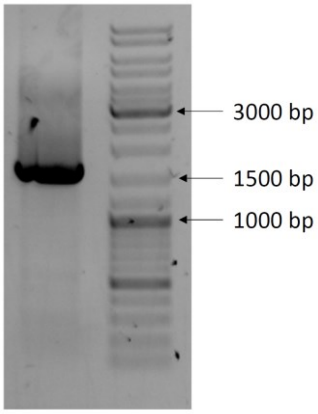


Figure IV-22. Amplification of the full-length HpCYP81AA2 cds (locus 928) from *H. perforatu*m cDNA using a proofreading polymerase.

10.2. Enzymatic activities of HpCYP81AA2 (locus 928)

Incubation of the microsomes containing HpCYP81AA2 with 2,3′,4,6-tetraHB resulted in the formation of one major and two minor products (**Figure IV-23a**). The major one with R_t 20.9 min represented 79.8 ± 1.1 % of the total product amount and was identified as 1,3,5-trihydroxyxanthone (1,3,5-triHX) by its UV spectrum, which showed three absorption maxima at 217, 245 and 313 nm matching those of authentic 1,3,5-triHX (**Figure IV-23b**). In addition, its MS spectrum showed a molecular ion peak at m/z 245 equivalent to $[M + H]^+$ of trihydroxylated xanthones (**Figure IV-23c**). The fragmentation pattern agreed with that of the reference compound.

The first minor product with R_t 19.6 min represented 11.0 ± 0.2 % of the total product amount and was identified as 1,3,7-triHX based on its UV and MS spectra. Another minor product with R_t 21.3 min represented 9.3 ± 0.9 % of the total enzymatic product. Its UV spectrum showed three absorption maxima at 236, 263 and 324 nm (**Figure IV-23b**) and did not match any spectrum of the available reference benzophenones and xanthones. ESI-MS could not identify the molecular mass of this compound which, therefore, was named unidentified product 1 (UIP1).

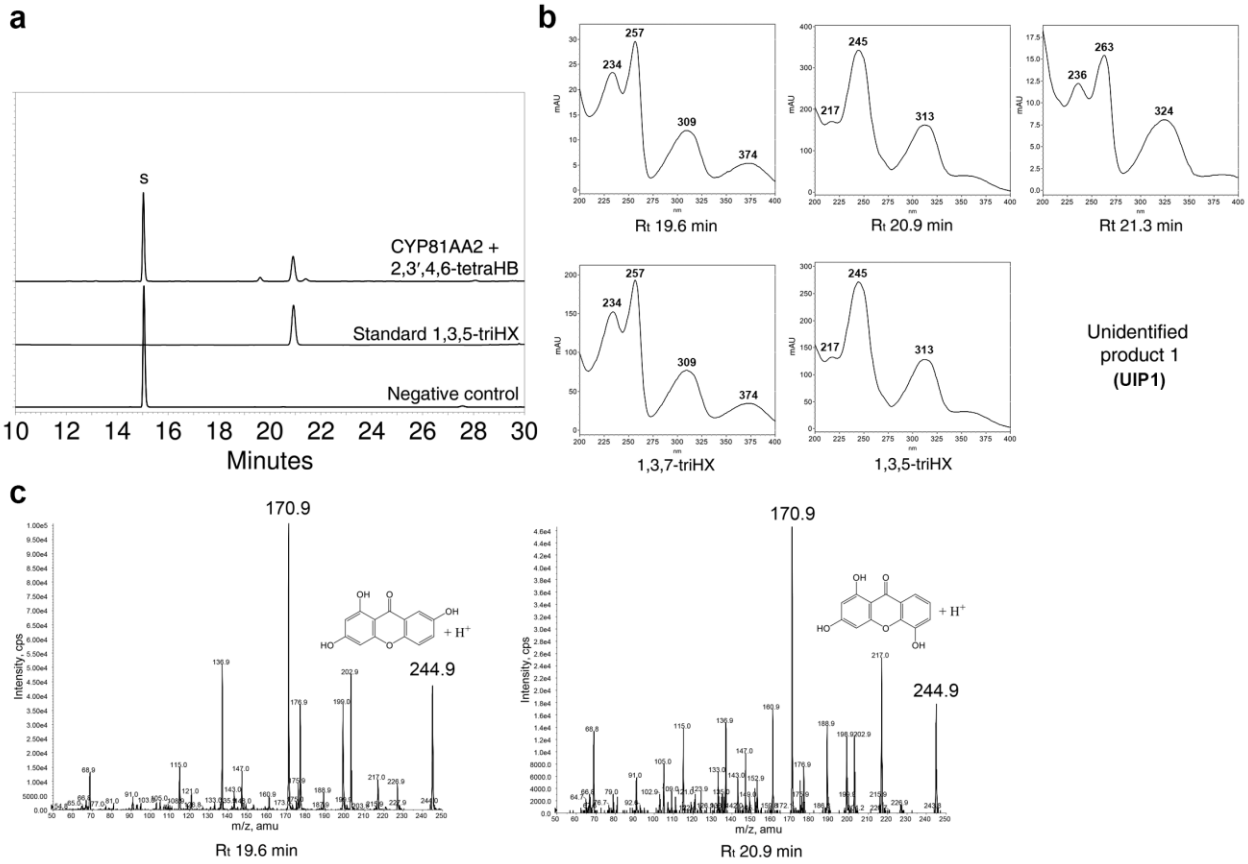


Figure IV-23. Incubation of the microsomes containing HcCYP81AA2 and HcCPR2 with 2,3′,4,6-tetraHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standard and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic products and standards. (**c**) EPI mass spectra of the enzymatic products.

Incubation with 2,4,6-triHB led to the formation of two major and two minor products (**Figure IV-24a**). The major product with R_t 20.9 min was identified as 1,3,5-triHX by UV and MS data as described in the previous incubation. The other major product with R_t 15.1 min was identified as 2,3′,4,6-tetraHB. Its UV spectrum matched that of authentic 2,3′,4,6-tetraHB and showed absorption maxima at 205, 258 and 305 nm (**Figure IV-24b**). EPI-MS showed a molecular ion peak at m/z 247 equivalent to $[M + H]^+$ of tetrahydroxylated benzophenones, in addition to two other peaks (m/z 121 and 153), which correspond to the ions resulting from breaking the bond at either side of the carbonyl group of the benzophenone skeleton (**Figure IV-24c**).

Similar to the previous incubation, the first minor product with R_t 19.6 min was identified as 1,3,7-triHX based on its UV and MS spectra (**Figure IV-24b,c**). The other minor product with R_t 21.3 min could not be identified but had a similar retention time and UV spectrum as the unidentified product 1 (UIP1), detected in the incubation with 2,3′,4,6-tetraHB.

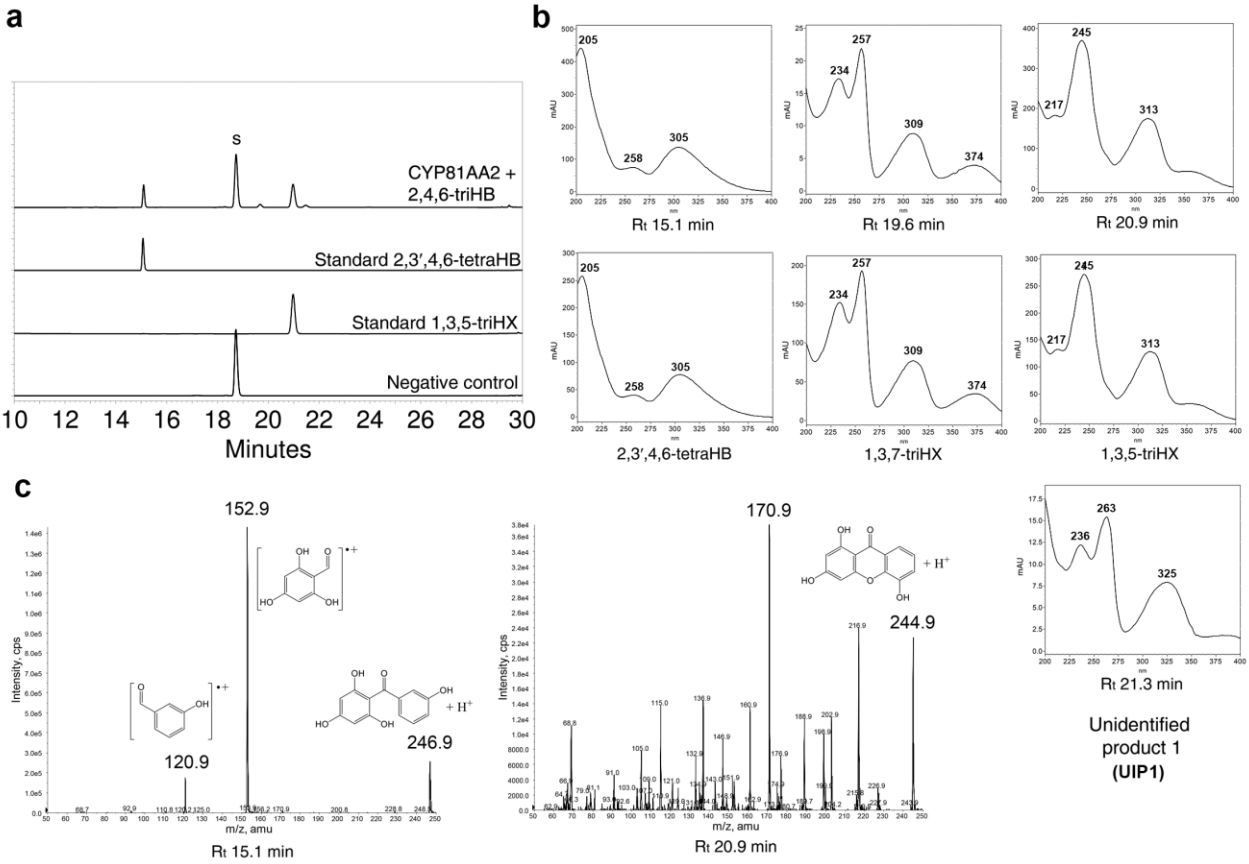


Figure IV-24. Incubation of the microsomes containing HcCYP81AA2 and HcCPR2 with 2,4,6-triHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standards and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic products and standards. (**c**) EPI mass spectra of the enzymatic products.

Incubation with 2,4-diHB yielded one major and two minor products (**Figure IV-25a**). The major product with R_t 15.1 min was identified as 2,2′,3,4′-tetrahydroxybenzophenone (2,2′,3,4′-tetraHB) by its UV and MS spectra. The UV spectrum matched that of authentic 2,2′,3,4′-tetraHB and showed absorption maxima at 218, 285 and 319 nm (**Figure IV-25b**). EPI-MS showed a molecular ion peak at m/z 247 equivalent to $[M + H]^+$ of tetrahydroxylated benzophenones. The peak at m/z 137 corresponds to the ions resulting from breaking the bond at either side of the carbonyl group, thereby confirming the presence of two hydroxyl groups on each aromatic ring of the benzophenone skeleton (**Figure IV-25c**). Similar to the incubation of CYP81AA1 with 2,4-diHB, the R_t 17.1 product was identified as 2,3′,4-triHB by MS. The peaks at m/z 137 and 121 confirmed the presence of mono- and dihydroxylated aromatic rings of the benzophenone skeleton (**Figure IV-25c**). The product with R_t 16.2 min showed UV absorption maxima at 234, 264, 313 and 376 nm which did not match any data of the available benzophenone and xanthone reference compounds. Therefore, xanthone cyclization products of 2,2′,3,4′- and 2,2′,4,5′-tetraHB can be ruled out. Again, ESI-MS failed to identify its molecular mass and it was named unidentified product 2 (UIP2).

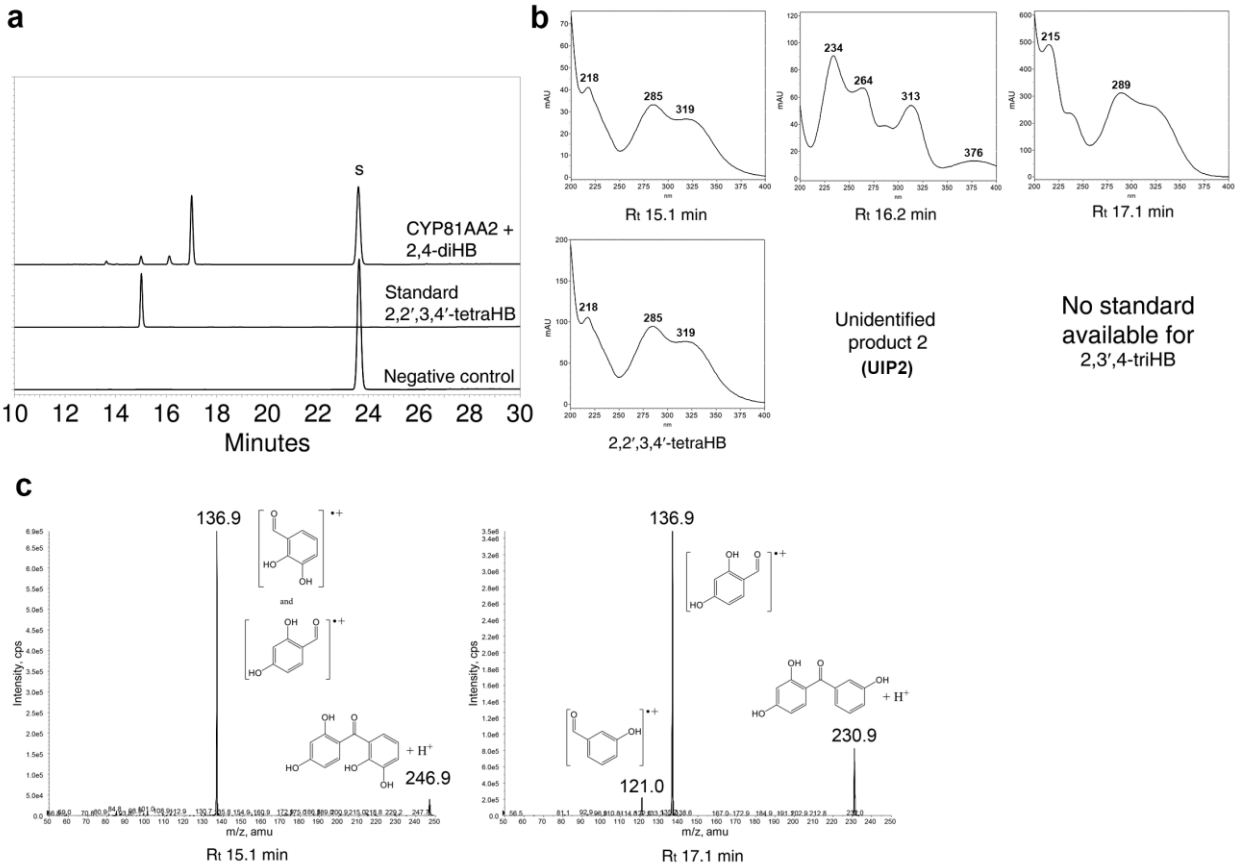


Figure IV-25. Incubation of the microsomes containing HcCYP81AA2 and HcCPR2 with 2,4-diHB. (**a**) Stacked HPLC-DAD chromatograms for the incubation, the available standard and the negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic products and available standard. (**c**) EPI mass spectra of the enzymatic products.

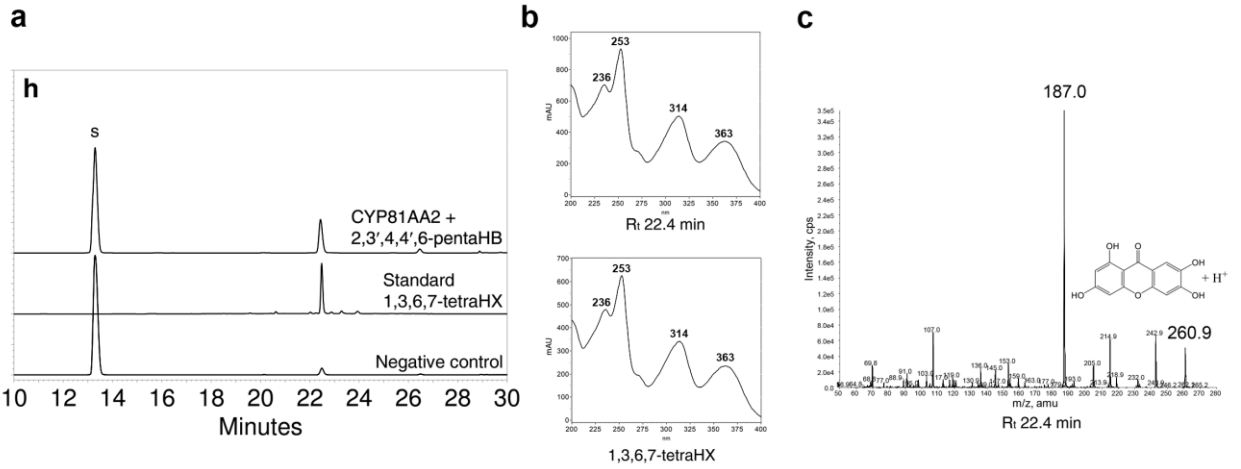


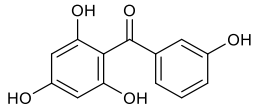
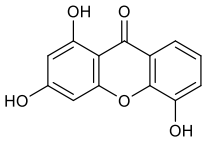
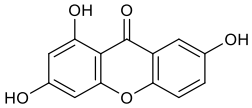
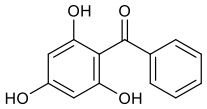
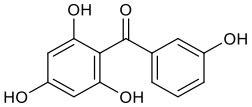
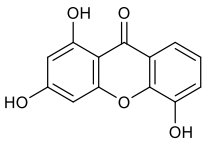
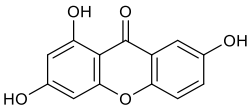
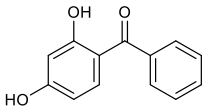
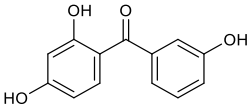
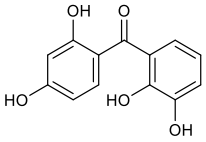
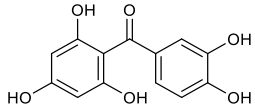
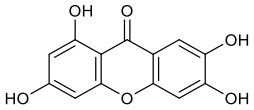
Figure IV-26. Incubation of the microsomes containing HcCYP81AA2 and HcCPR2 with 2,3′,4,4′,6-pentaHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standard and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic product and standard. (**c**) EPI mass spectrum of the enzymatic product.

Incubation with 2,3′,4,4′,6-pentaHB (maclurin) yielded a single product (R_t 22.4 min) (**Figure IV-26a**). This product was identified as 1,3,6,7-tetraHX (norathyriol) by its UV and MS spectra (**Figure IV-26b,c**). No trace of the expected 1,3,5,6-tetraHX could be detected. Again, traces of 1,3,6,7-tetraHX as a natural contaminant in 2,3′,4,4′,6-pentaHB samples were present in the negative control (Jefferson and Scheinmann, 1965).

Negative controls lacking NADPH or containing microsomes of yeast cells harboring the empty plasmid failed to produce any of the previous products. No activity was detectable with 2-HB, 4-HB, 1,3-diHX, 1,3,7-triHX, 1,3,5-triHX, 1,3,5,6-tetraHX and 1,3,6,7-tetraHX.

In conclusion, activity was detected with four benzophenones of the eleven substrates tested (**Table IV-5**). Accordingly, HpCYP81AA2 was identified as a bifunctional enzyme possessing both B3′H and 1,3,5-trihydroxyxanthone synthase (1,3,5-TXS) activities.

Table IV-5. Summary of the enzymatic activities of HpCYP81AA2.

Substrate	Product(s)			
 2,3′,4,6-tetraHB	 1,3,5-triHX	 1,3,7-triHX (minor)	UIP1 (minor)	
 2,4,6-triHB	 2,3′,4,6-tetraHB	 1,3,5-triHX	 1,3,7-triHX (minor)	UIP1 (minor)
 2,4-diHB	 2,3′,4-triHB	 2,2′,3,4′-tetraHB	UIP2 (minor)	
 2,3′,4,4′,6-pentaHB	 1,3,6,7-tetraHX			
2-HB	No product detected			
4-HB	No product detected			
1,3-diHX	No product detected			
1,3,7-triHX	No product detected			
1,3,5-triHX	No product detected			
1,3,6,7-tetraHX	No product detected			
1,3,5,6-tetraHX	No product detected			

10.3. Improvement of incubation conditions

10.3.1. Effect of pH and temperature

The optimum pH was 7.0 with loss of 18.0 and 6.5 % activity at pH 6.5 and 7.5, respectively (**Figure IV-27a**). The optimum temperature was 30 °C with loss of 19.6 and 17.1 % activity at 25 and 35 °C, respectively (**Figure IV-27b**).

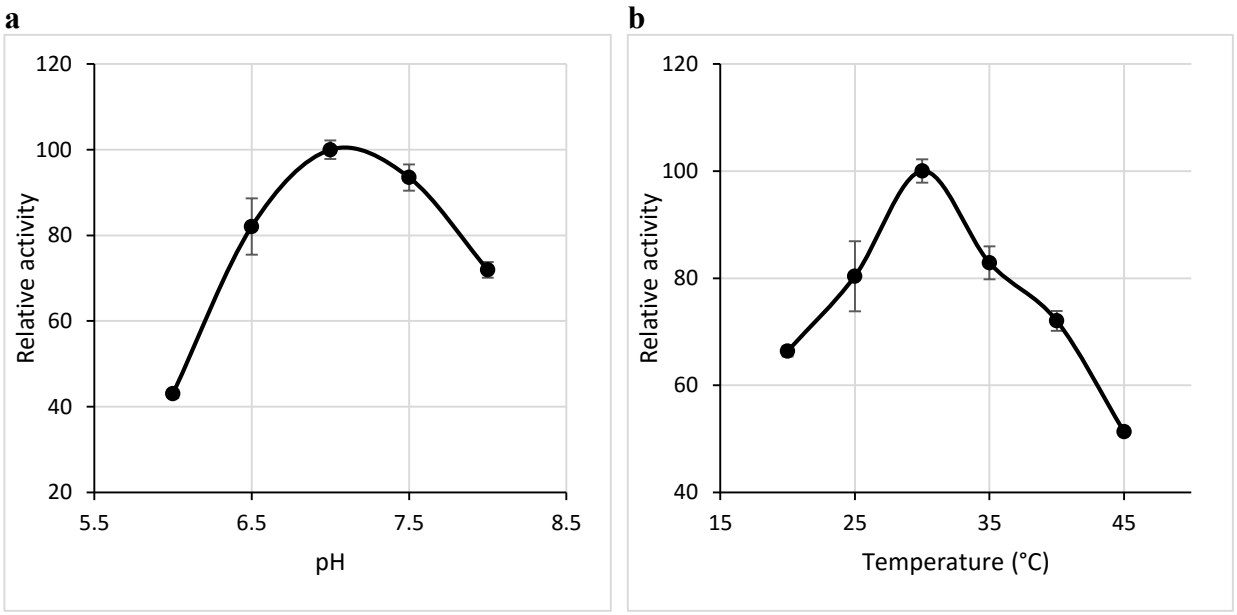


Figure IV-27. Effect of pH and incubation temperature on the activity of HpCYP81AA2. Data represent means ± s.d. (n = 3).

10.3.2. Effect of incubation time and protein amount

The accumulation of 1,3,5-triHX after incubation with 2,3′,4,6-tetraHB increased over the tested period of 120 min. The product formation rate was linear in the first 30 min of incubation (**Figure IV-28a**). For the protein concentration, product accumulation was linear up to approximately 50 µg microsomal protein in the standard assay (**Figure IV-28b**).

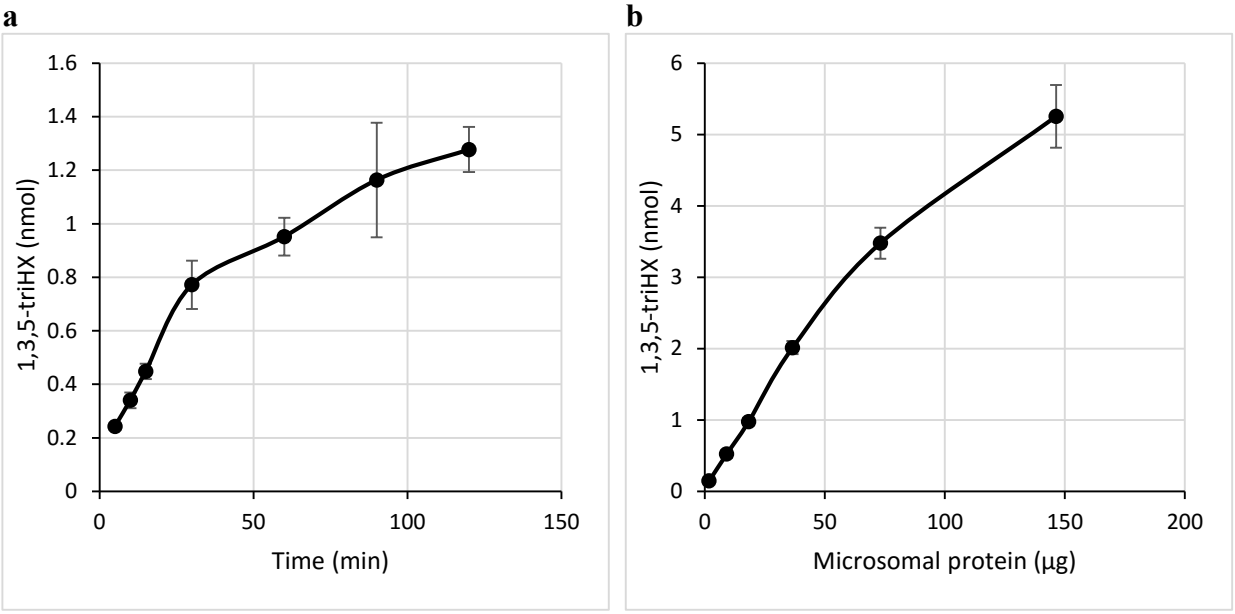


Figure IV-28. Effect of incubation time and protein amount on the activity of HpCYP81AA2. Data represent means ± s.d. (n = 3).

11. Functional characterization of locus 8128 from *H. perforatum* transcriptome

Primers with suitable restriction sites were designed according to the sequence of 'hpa_locus_8128_iso_3_len_1785_ver_2' from the MPGR databank. The cDNA obtained from *H. perforatum* mid-aged leaves was used as a template to amplify the ORF of locus 8128 with a proofreading polymerase. It was subsequently cloned between *Spe*I and *Pac*I restriction sites in MCS1 of the pESC-URA:HcCPR2 plasmid (III.2.3). The enzyme encoded by the ORF of locus 8128 was named HpCYP81AA3 (Nelson, 2006). The constructed expression plasmid (pESC-URA:HpCYP81AA3/HcCPR2) was transferred into INV*Sc*1 yeast cells (III.3.3.3), co-expression was induced and the microsomal fraction was isolated (III.4.2) and used for enzymatic incubations with the above list of potential substrates (**Figure IV-11**). The only reaction that was catalyzed by this enzyme was the formation of 1,3,6,7-tetraHX from 2,3′,4,4′,6-pentaHB (**Figure IV-29**). Neither B3′H nor TXS activity were detectable.

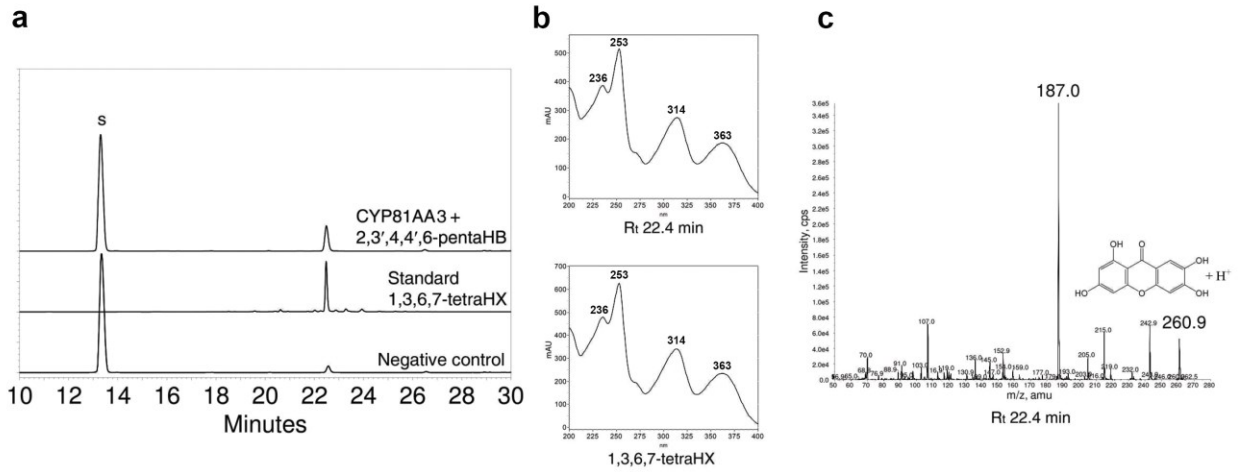


Figure IV-29. Incubation of the microsomes containing HcCYP81AA3 and HcCPR2 with 2,3′,4,4′,6-pentaHB. (**a**) Stacked HPLC-DAD chromatograms for incubation, standard and negative control; S stands for the substrate peak. (**b**) UV spectra of the enzymatic product and standard. (**c**) EPI mass spectrum of the enzymatic product.

12. Selection of targets for reciprocal mutagenesis

To gain insight into the regioselectivity-determining elements of CYP81AA1 and CYP81AA2, their amino acid sequences were aligned, secondary structure elements were predicted (Sirim et al., 2010), residues were numbered according to the standardized numbering scheme for class II CYPs (Gricman et al., 2014), and the putative six substrate recognition sites (SRSs) were identified (Gotoh, 1992). The predicted secondary structure elements of the two enzymes matched, except for β3-3, β4-1 and β4-2 located toward the C-termini (**Figure IV-30**). Therefore, the 55 C-terminal amino acids were reciprocally exchanged between the two enzymes (III.2.9) as a first round of mutations.

The second round involved reciprocal site-directed mutagenesis of individual amino acids (III.2.10). The two CYP81AA1 orthologs differed in only one residue (V/I220, standard position 189) within the six predicted SRSs. In contrast, CYP81AA1 and CYP81AA2 had 15 different residues within the SRSs, some of which were expected to be responsible for the alternative regioselectivities (**Figure IV-30**). Thus, changing one or more of these residues toward their counterparts in the other enzyme was likely to affect the nature of the cyclization product. To reduce the number of possible single and multiple reciprocal mutations to be studied, homology models of both enzymes were built (III.6.2), based on the structure of the closed conformation of mammalian CYP2B4 in complex with ticlopidine (3KW4) (Gay et al., 2010). Examination of the superimposed models revealed that 17 residues were located within a 4 Å radius of the bound inhibitor in either protein model. Five of these residues belonging to the SRSs 5 and 6 differed between the two enzymes (**Table IV-6, Figure IV-31**). In addition, S375 in CYP81AA1, which corresponds to L378 in CYP81AA2 (standard position 330.1), was selected as a mutagenesis target because this position was previously predicted to affect the selectivity of CYPs (Seifert and Pleiss, 2009) and undergoes a change from hydrophilic to hydrophobic.

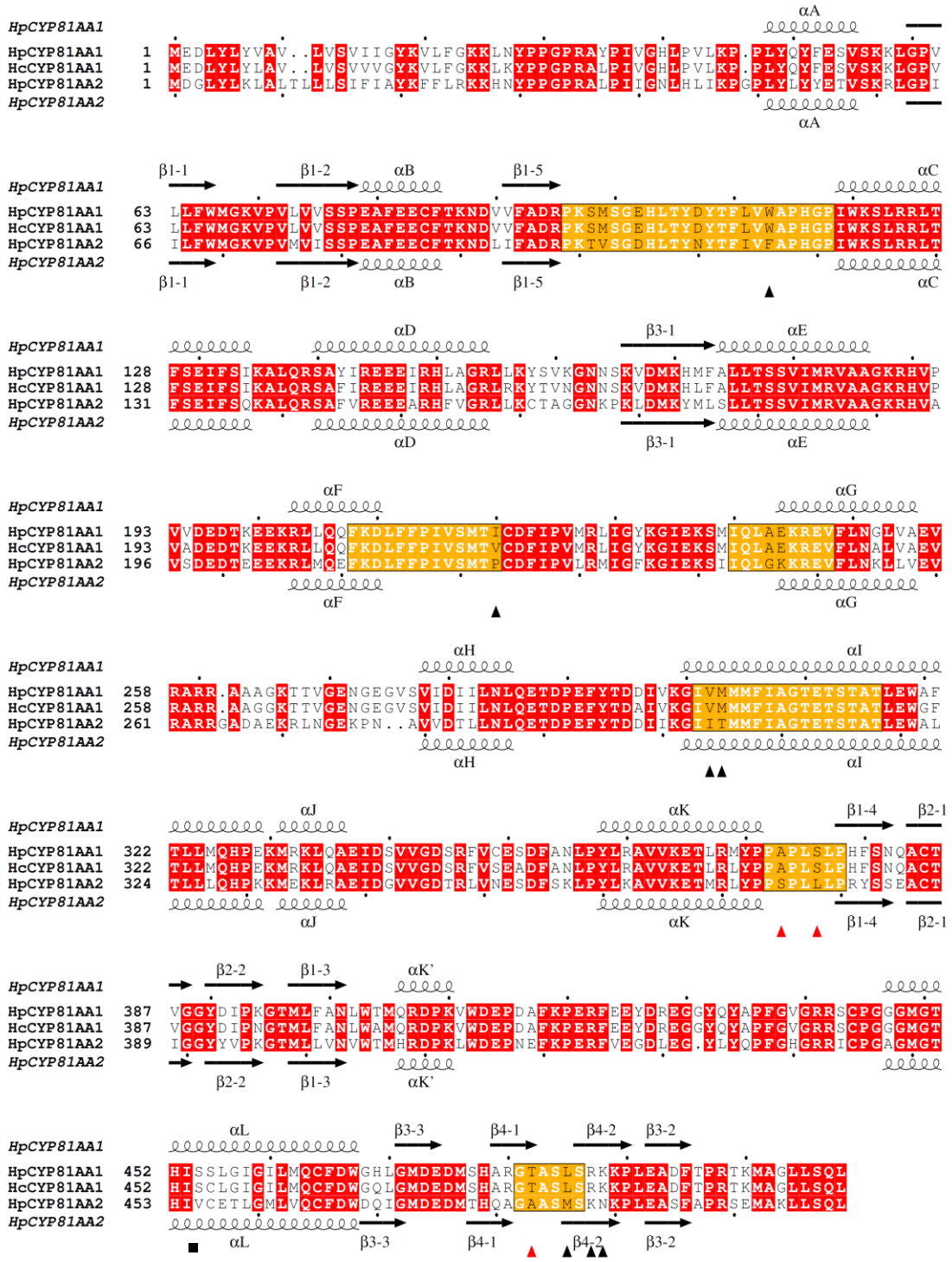


Figure IV-30. Multiple sequence alignment of HpCYP81AA1, HcCYP81AA1 and HpCYP81AA2. Amino acids expected to be present in SRSs 1-6 are highlighted by orange boxes. Dissimilar residues are written in black letters and conserved residues outside the SRSs are represented in white with a red background. The conserved secondary structures of HpCYP81AA1 and HpCYP81AA2 are indicated at the top and the bottom, respectively. The black triangles point to positions selected for mutation and the red triangles refer to the major regioselectivity-determining residues in HpCYP81AA2. The black square refers to the position of the C-terminal exchange.

Table IV-6. Contact residues located within a 4 Å radius of the bound inhibitor and their standard positions and SRS numbers. Divergent residues are underlined.

CYP81AA1		CYP81AA2		Standard numbering position	SRS number
Residue	Position	Residue	Position		
PHE	110	PHE	113	85	1
PHE	212	PHE	215	181	2
ILE	215	ILE	218	184	2
VAL	216	VAL	219	185	2
ILE	307	ILE	309	263	4
ALA	308	ALA	310	264	4
GLU	311	GLU	313	267	4
THR	312	THR	314	268	4
ALA	315	ALA	317	271	4
TYR	370	TYR	372	325	5
PRO	372	PRO	374	327	5
ALA	373	SER	375	328	5
LEU	377	LEU	379	331	5
THR	482	MET	486	437/439	6
ALA	483	SER	487	437.3/440	6
SER	484	LYS	488	438/441	6
LEU	485	ASN	489	439/441.1	6

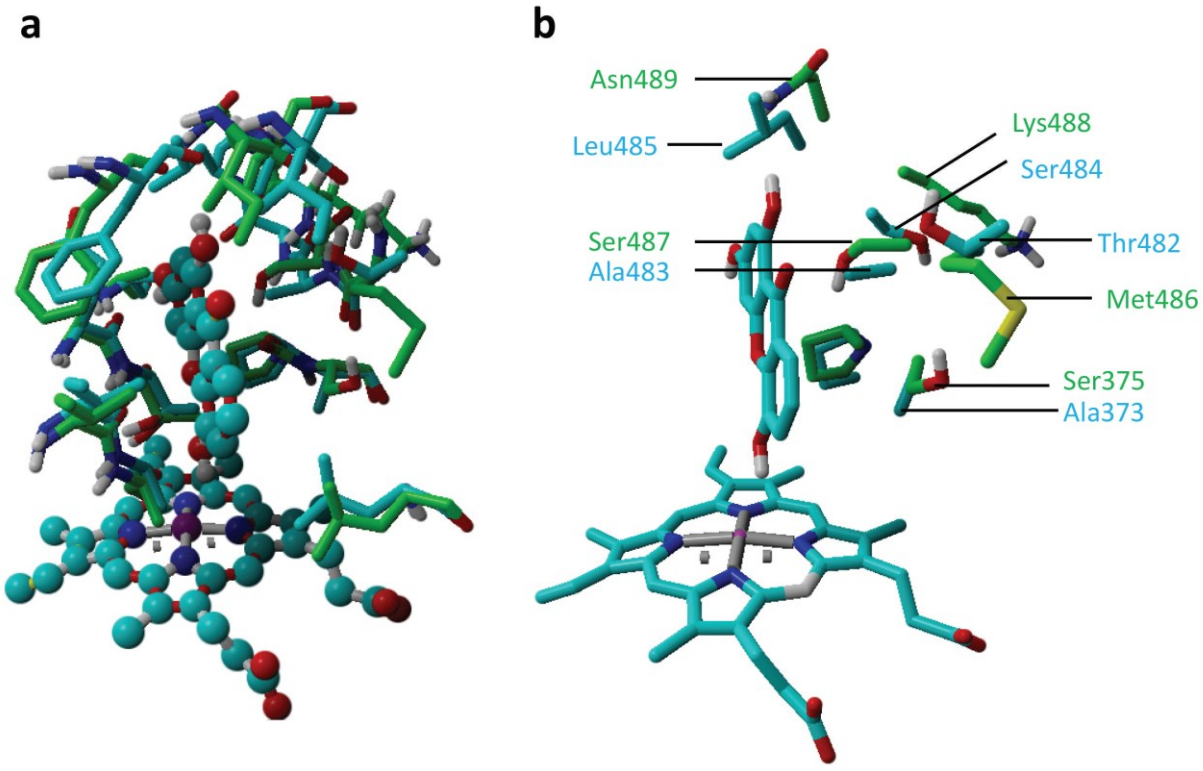


Figure IV-31. Superimposed models of CYP81AA1 (cyan) and CYP81AA2 (green) showing (**a**) all contact residues within a 4 Å radius of the product and (**b**) the divergent residues.

13. Control of the 1,3,5-TXS (CYP81AA2) regioselectivity

Wild-type CYP81AA2 released 79.5 % of the total product as 1,3,5-triHX and 11.0 % as 1,3,7-triHX. Reciprocal mutation of six individual residues in SRSs 5 and 6 toward their counterparts in CYP81AA1 led to a gradual change in the 1,3,5-triHX to 1,3,7-triHX ratio (**Figure IV-32**). In the homology model built, the standard positions 437 and 328 were the closest divergent contact residues of the bound inhibitor. Reciprocal single substitutions at the positions 437 in SRS-6 (A483T) and 328 in SRS-5 (S375A) increased the 1,3,7-triHX portion from 11.0 to 19.2 and 30.2 %, respectively. Single substitution at the selected standard position 330.1 in SRS-5 (L376S) yielded only 16.7 % 1,3,7-triHX. Of the double mutants generated, L378S/A483T formed 22.9 % 1,3,7-triHX, however, S375A/L378S and S375A/A483T, which contained the standard position 328, resulted in 50.5 and 59.3 % 1,3,7-triHX, respectively. This outcome was increased to 80.7 % by combining all three positions in the S375A/L378S/A483T mutant. Finally, the sextuple mutant S375A/L378S/A483T/M486L/K488R/N489K (mut6), which involved all mismatching contact residues, released 91.1 % of the total product as 1,3,7-triHX (**Figure IV-32**). None of the mutants generated could mimic the absolute regiospecificity of CYP81AA1 toward formation of 1,3,7-triHX. No appreciable changes in the level of the unidentified product were observed.

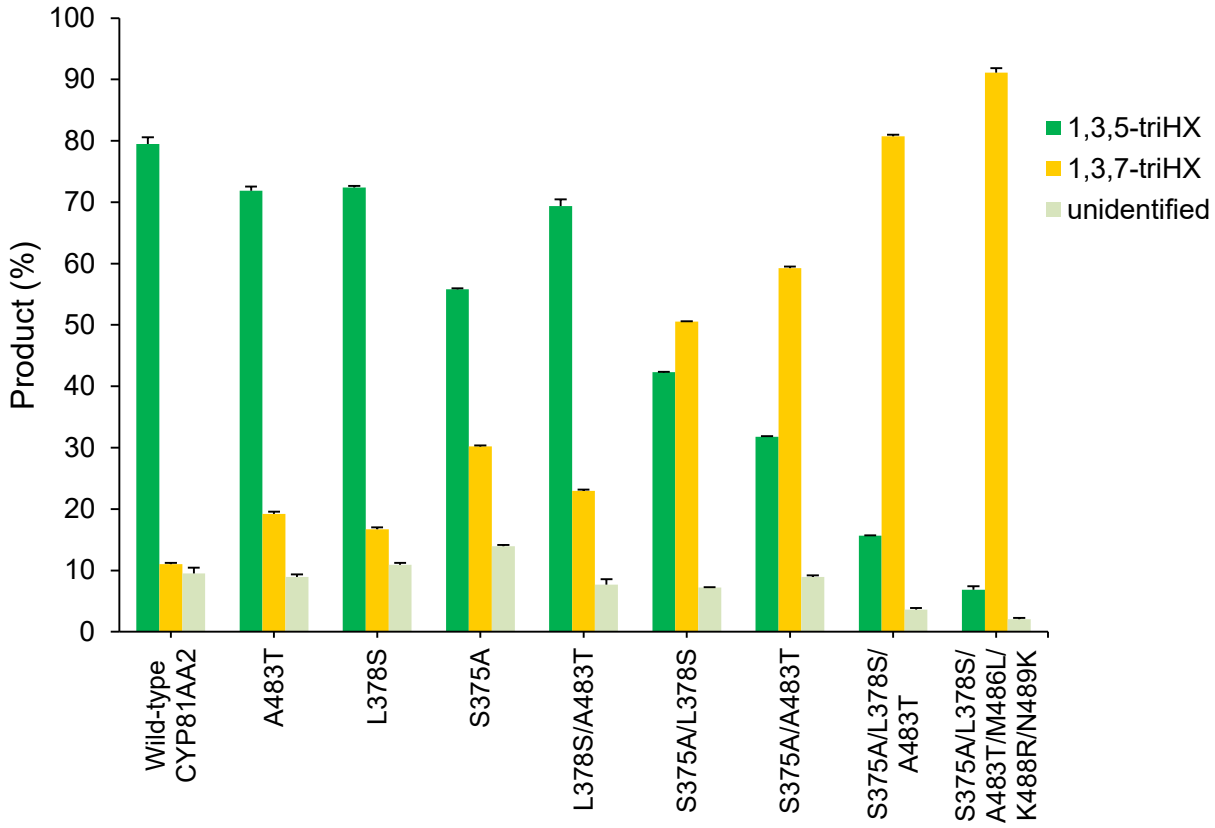


Figure IV-32. Product profiles of wild-type HpCYP81AA2 and mutant enzymes. Data represent means ± s.d. (n = 4).

14. Attempts toward changing the regiospecifity of CYP81AA1

Wild-type CYP81AA1 showed absolute specificity toward the production of 1,3,7-triHX. Substitution of the 55 C-terminal residues by those of CYP81AA2 failed to cause any change in the regiospecificity. Similarly, reciprocal mutations in SRSs 5 and 6 had only minor effects on the regioselectivity (**Table IV-7**). For example, the sextuple mutation A373S/S376L/T482A/L485M/ R487K/K488N caused only 6.5 % of the total product to be released as 1,3,5-triHX. This result resembled the effect of the single mutant A373S, which formed 5.1 % 1,3,5-triHX. Replacing alanine in the T482A position of the triple mutant (A373S/S376L/T482A) with larger hydrophobic residues (T482V and T482F) failed to further shift the regioselectivity. Finally, a third set of mutations were created by selecting divergent residues within an increased radius (< 15 Å of the inhibitor ticlopidine), such as W113F in SRS-1, I220P in SRS-2, and V302I/M303T in SRS-4. However, none of the additional mutations in SRSs 1 and 2 led to further changes in the regioselectivity. Mutations performed on SRS-4 even resulted in the loss of enzyme activity.

Table IV-7. Product profiles of HpCYP81AA1 mutants

Enzyme mutant	1,3,7-triHX (% of total product)	1,3,5-triHX (% of total product)
Wild-type CYP81AA1	100	0
CYP81AA1-AA2	100	0
CYP81AA1 A373S	94.9 ± 1.1	5.1 ± 1.1
CYP81AA1-AA2 A373S	93.2 ± 0.2	6.8 ± 0.2
CYP81AA1 A373S/S376L	95.0 ± 0.4	5.0 ± 0.4
CYP81AA1-AA2 A373S/S376L	96.0 ± 0.8	4.0 ± 0.8
CYP81AA1 A373S/S376L/T482A	95.5 ± 0.3	4.5 ± 0.3
CYP81AA1 A373S/S376L/T482A/L485M/R487K/K488N	93.5 ± 0.6	6.5 ± 0.6
CYP81AA1 A373S/S376L/T482V	94.6 ± 0.5	5.4 ± 0.5
CYP81AA1 A373S/S376L/T482F	95.9 ± 0.3	4.1 ± 0.3
CYP81AA1 W113F/A373S/S376L/T482A	95.2 ± 0.1	4.8 ± 0.1
CYP81AA1 I220P/A373S/S376L/T482A	99.4 ± 0.1	0.6 ± 0.1
CYP81AA1 A373S/S376L/T482A/L485M/R487K/K488N/V302I/M303T	No activity detected	
CYP81AA1 W113F/I220P/A373S/S376L/T482A/L485M/R487K/K488N/ V302I/M303T	No activity detected	

15. Product docking

Docking of 1,3,5- and 1,3,7-triHX into the homology models of wild-type CYP81AA2 and its sextuple mutant S375A/L378S/A483T/M486L/K488R/N489K (mut6) revealed different modes of product binding, indicating hindrance of internal rotation in the 2,3′,4,6-tetraHB substrate (**Figure IV-33**). In the wild-type enzyme (1,3,5-TXS), the 3-hydroxy group of the product can form a hydrogen bond with S375, orienting it toward the heme loop. The 5-hydroxy group points

toward the heme group. At the substrate level, rotation of the 3′-hydroxyphenyl ring of 2,3′,4,6-tetraHB is hindered by the side chains of the I-helix residues. In the sextuple mutant, the 3-hydroxy group of the product can form a new hydrogen bond with A483T, orienting it toward the C-terminal loop. This hydrogen bond switching can explain the observed cooperative effect of the double mutant S375A/A438T. The mutation introduced in the former H-bonded residue S375A together with the nearby mutation L378S may alter the conformation of the heme loop to cooperatively support the reorientation of the substrate. The 7-hydroxy group of the product is located between the I-helix and the heme group, which again indicates hindrance of free rotation of the 3′-hydroxyphenyl ring of 2,3′,4,6-tetraHB.

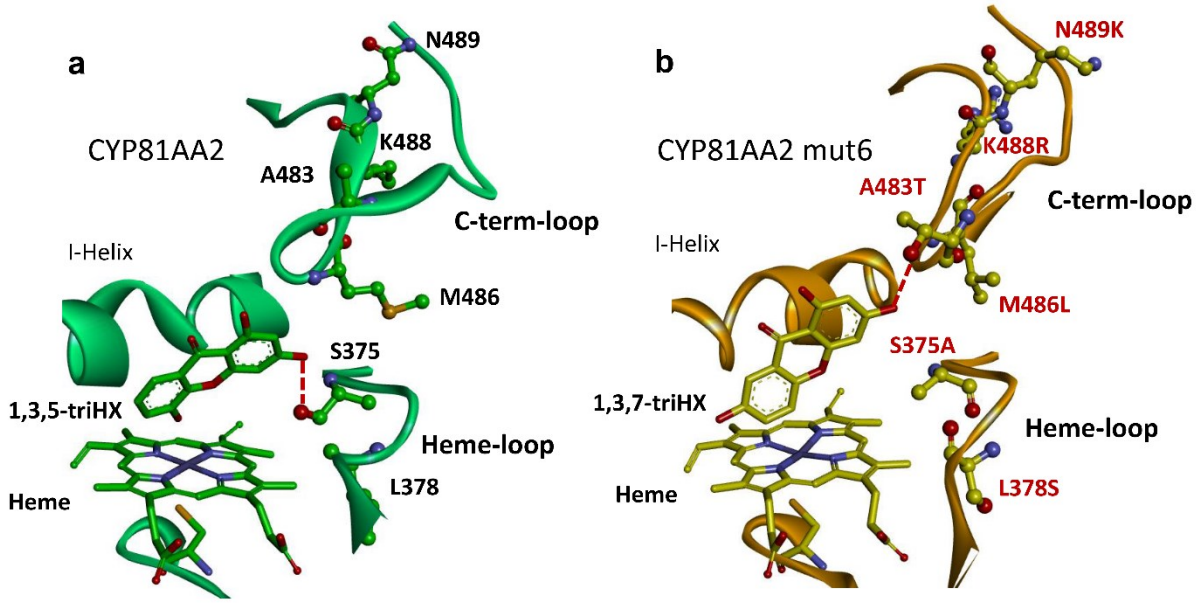


Figure IV-33. Product docking. (**a**) Active site of wild-type CYP81AA2 (lime) with bound 1,3,5-triHX (green sticks). The 3-hydroxy group forms a hydrogen bond with S375, orienting it toward the heme loop. (**b**) Active site of the sextuple mutant (mut6) of CYP81AA2 (orange) with bound 1,3,7-triHX (yellow sticks). The 3-hydroxy group forms a hydrogen bond with A483T, orienting it toward the C-term-loop.

V. Discussion

Plants produce a variety of secondary metabolites which play a major role in their defense against pathogens and herbivores. In addition, many of those metabolites possess valuable pharmacological activities, beneficial to human health and therefore serve as potential lead compounds for the production of new drugs. The biosynthesis of secondary metabolites in plants involves series of reactions catalyzed by various classes of enzymes which direct the metabolic flux towards the production of complex molecules from relatively simple primary building blocks. Deciphering the biosynthetic methodology, by which plants synthesize those products, and studying the involved enzymes would enable us utilizing them in metabolic engineering applications for the heterologous production of natural products (Renault et al., 2014). In addition, overexpression of some genes and suppression of others may lead to generation of transgenic plants with customized metabolite profiles.

Among the enzymes responsible for biosynthesis of natural products are cytochrome P450 enzymes (CYPs). They constitute a large superfamily of heme-thiolate proteins found in almost all living organisms. CYPs catalyze regio- and stereoselective oxidative attack of non-activated carbons at physiological conditions (Werck-Reichhart and Feyereisen, 2000). In plants, they are encoded by around 1% of the protein-coding genes and the catalytic functions for most of their gene products remain unknown (Nelson and Werck-Reichhart, 2011). Assigning functions to new CYPs is a challenge because of their large numbers, lack of reliable sequence-function correlations and the structural similarities between family members (Mizutani and Ohta, 2010). They are responsible for various reactions in plant primary as well as secondary metabolism. Generally, plant CYPs catalyze hydroxylation (monooxygenation) reactions. However, few enzymes exhibit less common unusual activities (Mizutani and Sato, 2011; Guengerich and Munro, 2013), including methylenedioxy bridge formation (Ikezawa et al., 2003; Ono et al., 2006; Ikezawa et al., 2007), successive oxidation of a single carbon (Helliwell et al., 2001; Carelli et al., 2011), rearrangement of carbon skeletons (Akashi et al., 1999), lactone ring formation (Zhang et al., 2014), sterol desaturation (Morikawa et al., 2006; Field and Osbourn, 2008), S-heterocyclization (Klein and Sattely, 2015), C–C bond cleavage (Irmler et al., 2000), and phenol coupling (Kraus and Kutchan, 1995; Ikezawa et al., 2008; Gesell et al., 2009). Intramolecular C–O phenol couplings were biochemically established in plant xanthone biosynthesis (Peters et al., 1997). However, no genes encoding those enzymes have been isolated so far. Several CYP-catalyzed intramolecular phenol coupling reactions in plants, bacteria and humans have been extensively studied at the gene level as subsequently surveyed.

1. CYP-catalyzed intramolecular phenol coupling reactions

1.1. Glycopeptide antibiotic biosynthesis in bacteria

Crosslinking of aromatic residues by intramolecular phenol coupling reactions is critical in conferring the three-dimensional structure in glycopeptide antibiotic biosynthesis (Haslinger et al., 2014; Yim et al., 2014). Four CYPs (OxyA, B, C and E) catalyze intramolecular C–O and C–C phenol couplings during the biosynthesis of vancomycin and teicoplanin-type glycopeptide antibiotics (**Figure V-1**) (Bischoff et al., 2001; Cryle et al., 2011). However, those CYPs do not

act on the free substrates. Their heptapeptide substrates are covalently bound to an acyl carrier domain of the non-ribosomal peptide synthase (NRPS) (Woithe et al., 2007; Cryle et al., 2011). Recently, the involvement of an X-domain in the NRPS in the recruitment of those oxygenases has been established (Haslinger et al., 2015). The order of the coupling reactions has been identified (Bischoff et al., 2001; Stegmann et al., 2006) and crystal structures for some of the bacterial soluble proteins have been solved (Cryle et al., 2011; Li et al., 2011). However, the exact mechanism of the phenol coupling reaction in light of the catalytic cycle of CYPs has never been discussed in detail for those enzymes.

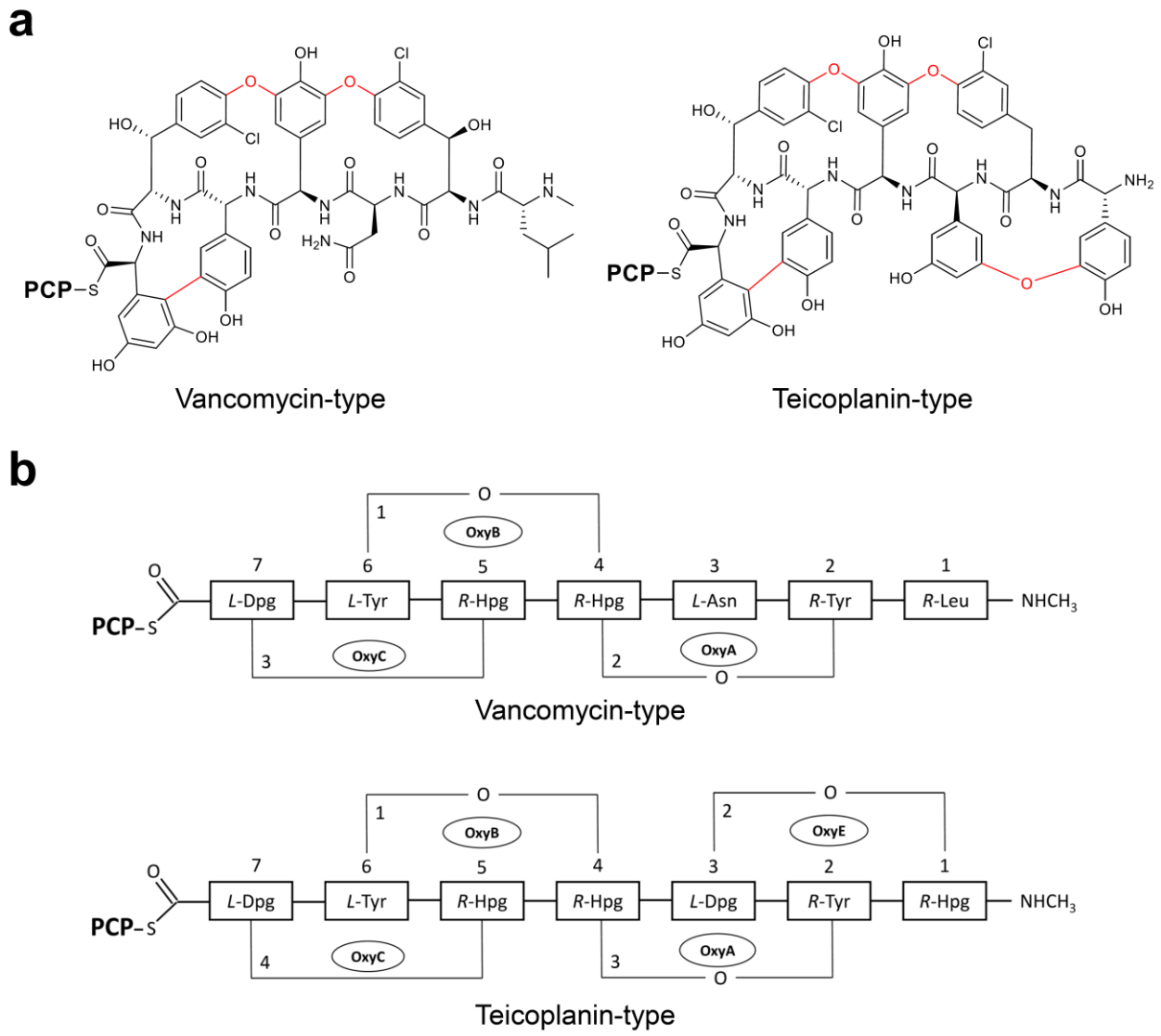


Figure V-1. Phenol couplings catalyzed by the Oxy CYPs in vancomycin and teicoplanin-type antibiotic biosynthesis. (**a**) Structure of the vancomycin and teicoplanin aglycone peptides bound to the peptidyl carrier protein (PCP), showing the bonds synthesized by the phenol coupling Oxy enzymes highlighted in red. (**b**) Schematic representation of the peptides sequences of vancomycin and teicoplanin, showing the order of cross-linking steps and the enzymes involved; adapted from Cryle et al. (2011).

1.2. Mycocyclosin biosynthesis in *Mycobacterium tuberculosis*

The *M. tuberculosis* genome contains 20 CYP-coding genes, of which CYP121 was found to be essential for the viability of the bacterium (McLean et al., 2008). Moreover, a correlation was established between the binding constants of azole antifungal drugs to CYP121 and the MIC values of these drugs, indicating that CYP121 is the main target of azoles *in vivo* (McLean et al., 2008). The catalytic function of CYP121 was elucidated by Belin et al. (2009). The enzyme was found to convert the substrate cyclodityrosine (cYY) into a product called mycocyclosine (mcyc) through intramolecular C–C phenol coupling reaction (**Figure V-2**). The coupling occurs between the two tyrosine aromatic rings in *ortho* position to the hydroxyl group of each tyrosine moiety (Belin et al., 2009).

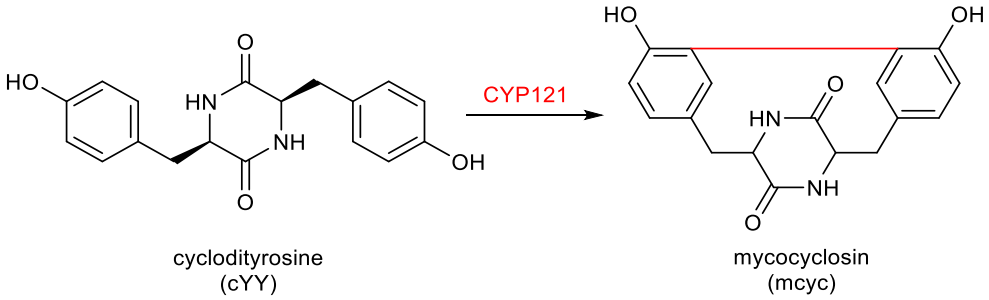


Figure V-2. Cyclodityrosine is converted via a C–C phenol coupling reaction to mycocyclosin by CYP121 in *M. tuberculosis*.

1.3. Isoquinoline alkaloid biosynthesis in plants

Intramolecular C–C and intermolecular C–O phenol coupling reactions were studied at the gene level in benzylisoquinoline alkaloid metabolism (Kraus and Kutchan, 1995; Ikezawa et al., 2008; Gesell et al., 2009). Such reactions are common in plant biosynthetic pathways and are often critical in the biosynthesis of plant secondary metabolites (Mizutani and Sato, 2011; Guengerich and Munro, 2013). To date, only three transcripts for plant CYPs catalyzing phenol coupling reactions have been cloned, all of which contribute to the biosynthesis of isoquinoline alkaloids. CYP80G2 and CYP719B1 catalyze intramolecular C–C couplings to convert (*S*)- and (*R*)-reticuline to (*S*)-corytuberine and salutaridine, respectively (**Figure V-3**) (Ikezawa et al., 2008; Gesell et al., 2009). CYP80A1 is the only known plant CYP that is able to catalyze an intermolecular C–O phenol coupling reaction between (*R*)- and (*S*)-*N*-methylcoclaurine to form the bisbenzylisoquinoline alkaloid berbamunine (**Figure V-3**) (Kraus and Kutchan, 1995). None of the previously characterized phenol coupling CYPs shows multifunctionality and no plant CYP catalyzing an intramolecular C–O phenol coupling reaction has been cloned previous to the present study.

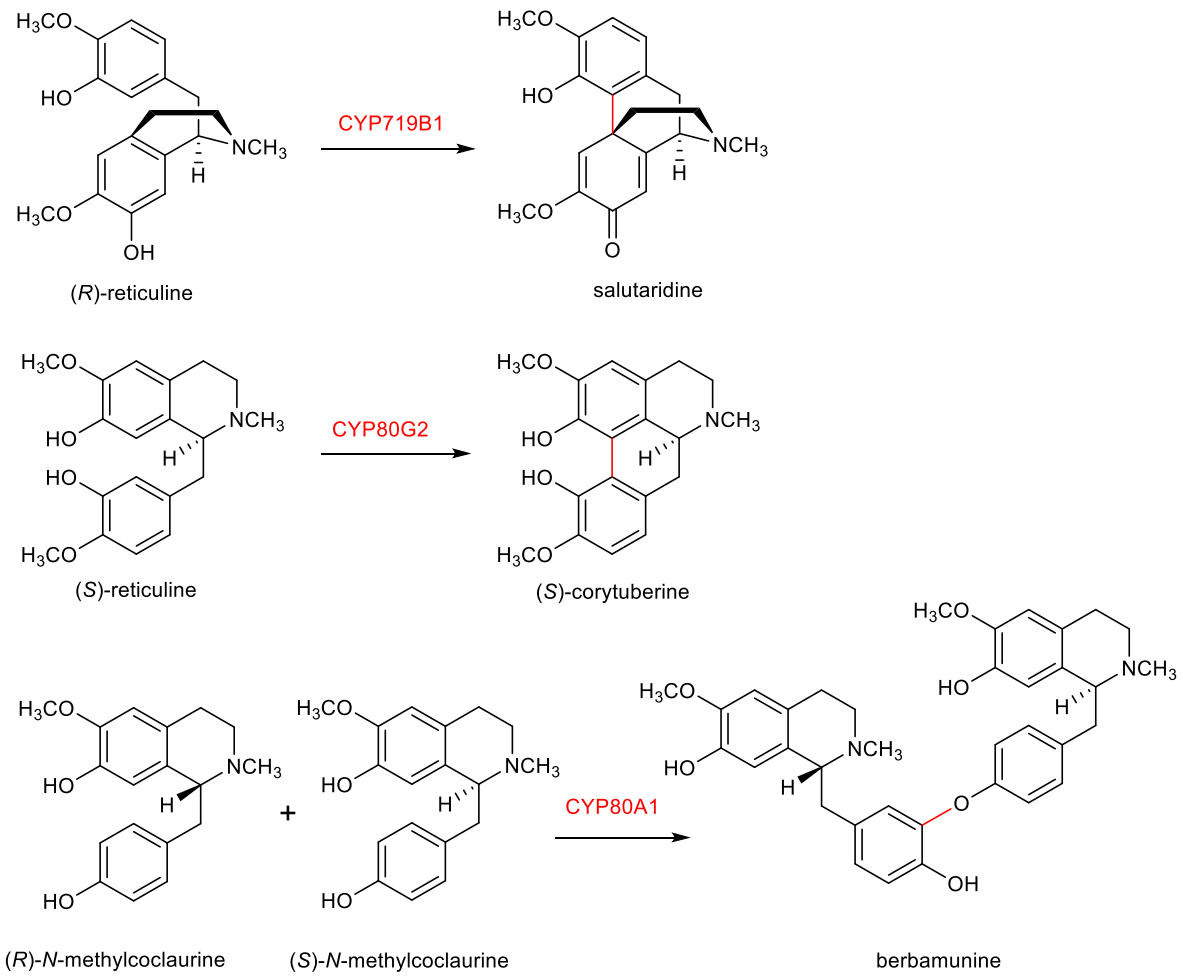


Figure V-3. Phenol coupling reactions catalyzed by plant CYPs in isoquinoline alkaloid biosynthesis.

1.4. Endogenous morphine biosynthesis in mammals

Analogous to plant salutaridine synthase (CYP719B1), human CYP2D4 and CYP3A4 were found to catalyze the phenol coupling of (*R*)-reticuline in the course of endogenous morphine biosynthesis. However, the reactions lacked the absolute regiospecificity observed for their plant counterpart, leading to the production of all four possible phenol coupling products (**Figure V-4**). A broad substrate spectrum and the lack of absolute regiospecificity are typical features of mammalian liver microsomal CYPs, as they are involved in xenobiotic detoxification (Grobe et al., 2009).

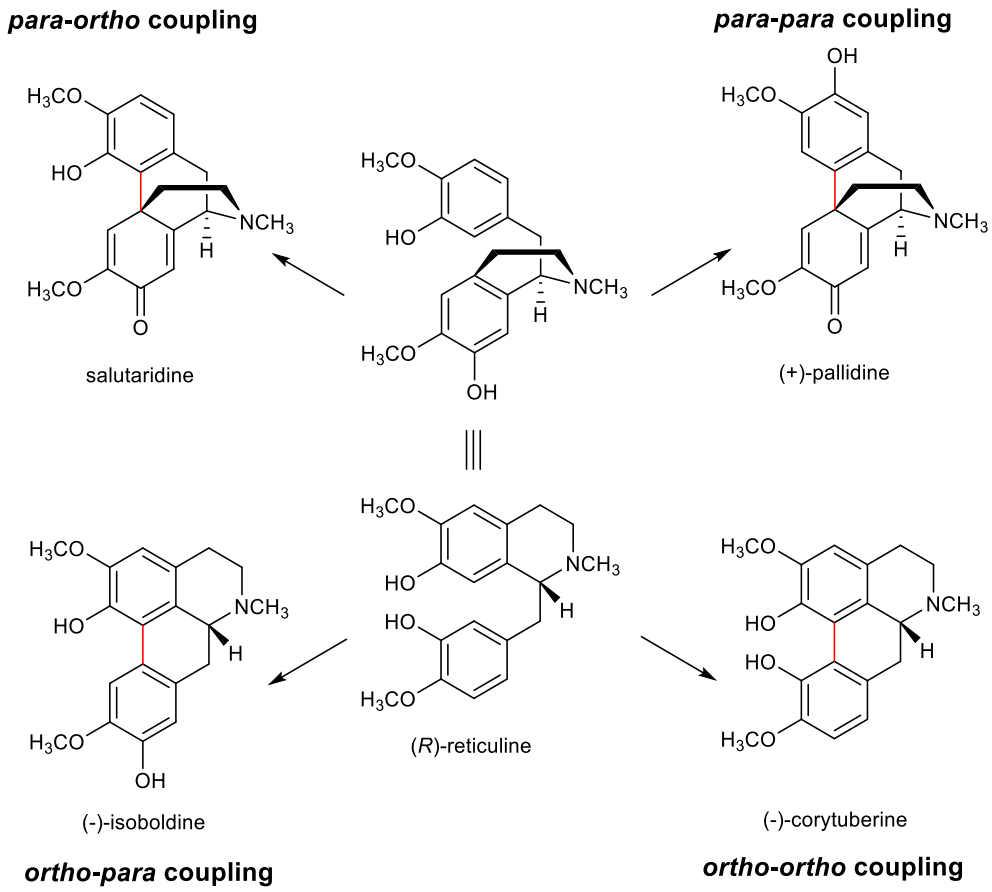


Figure V-4. Phenol coupling reactions in mammals. Human CYP2D6 and CYP3A4 catalyze the enzymatic conversion of (*R*)-reticuline into salutaridine, (+)-pallidine, (-)-isoboldine, and (-)-corytuberine by oxidative phenol coupling reactions; adapted from Grobe et al. (2009).

2. The role of CYP81AA subfamily members in plant xanthone biosynthesis

Xanthone (9*H*-xanthen-9-one) compounds constitute a group of natural products found in fungi, lichens and a few families of higher plants (Vieira and Kijjoa, 2005; El-Seedi et al., 2009; El-Seedi et al., 2010). Their distribution as well as oxygenation and prenylation patterns are considered of significant chemotaxonomic value (Demirkiran, 2007; Crockett and Robson, 2011). To date, more than 1,500 xanthone-based natural products are present in chemical databases. Plant xanthones function as defense compounds against herbivores and microorganisms (Franklin et al., 2009; Tocci et al., 2011). In humans, they exhibit a wide array of interesting pharmacological activities including anti-microbial, anti-protozoal, anti-HIV-1, anti-inflammatory, anti-atherosclerotic, cytotoxic and cancer chemopreventive activities (Demirkiran, 2007; El-Seedi et al., 2010). In addition, some xanthones possess a combination of acetylcholinesterase inhibitory, free radical scavenging and monoamine oxidase inhibitory activities. This combination of activities leads to growing interest in use for combating neurodegenerative diseases such as Alzheimer′s (Zhang, 2005; Martinez et al., 2012; Wang et al., 2012).

In plant xanthone biosynthesis, the interface between benzophenone and xanthone skeletons is a ring closure event, which converts the diphenyl ketone scaffold to a tricyclic ring system. This cyclization is preceded by a 3′-hydroxylation step, which favors subsequent substitution at the *ortho* and *para* positions. In the present work, it is demonstrated that both consecutive reactions are catalyzed by bifunctional CYPs, two variants of which (CYP81AA1 and CYP81AA2) accomplish the ring closure regioselectively (**Figure V-5**). Hp and HcCYP81AA1 catalyze the exclusive formation of 1,3,7-triHX as a final product either from 2,4,6-triHB or 2,3′,4,6-tetraHB and therefore received the name 1,3,7-trihydroxyxanthone synthase (1,3,7-TXS). On the other hand, HpCYP81AA2 accepted the same physiological substrates (2,4,6-triHB and 2,3′,4,6-tetraHB) to form a mixture of three final products, the major of which was 1,3,5-triHX and hence was named 1,3,5-trihydroxyxanthone synthase (1,3,5-TXS).

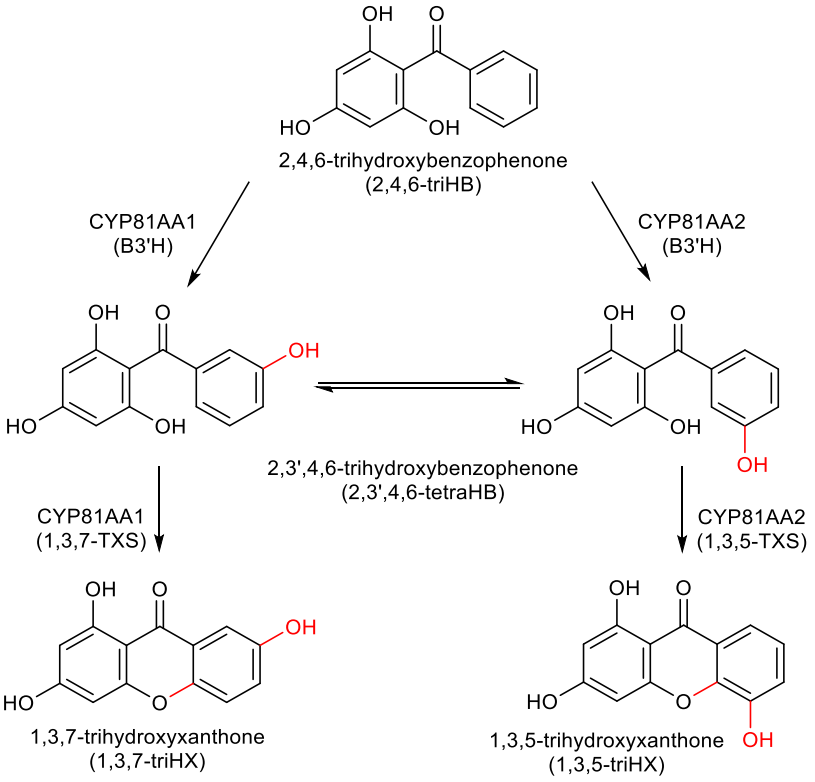


Figure V-5. The reactions catalyzed by CYP81AA1 and CYP81AA2 using the physiological substrates.

No hydroxylation at the xanthone level took place, as indicated by the non-acceptance of 1,3-dihydroxyxanthone and related compounds. Notably, 2,4-diHB was 3′-hydroxylated by both CYP81AA1 and CYP81AA2, however, the product failed to be cyclized because of the lack of the 6-hydroxy group and instead underwent 2′-hydroxylation, yielding 2,2′,4,5′- and 2,2′,3,4′-tetraHB, respectively. Subsequent loss of water, as previously proposed for *ortho-ortho′*-hydroxylated benzophenones (Kitanov and Nedialkov, 2001), was hindered by hydrogen bonding between the 2-hydroxy and carbonyl groups. With 2,3′,4,4′,6-pentaHB, the ring closure took place but lacked

regioselectivity, as both enzymes formed 1,3,6,7-tetraHX. This conversion was the only reaction catalyzed by CYP81AA3, whose physiological function remains open. None of the three CYPs converted monohydroxylated benzophenones, such as 2- and 4-hydroxybenzophenone.

The previously identified plant phenol coupling CYPs are members of CYP719 and CYP80 families, which are represented by a limited number of members in a restricted quantity of plant taxa (Bak et al., 2011). In contrast, CYP81 is a large enzyme family consisting of multiple members in all plant genomes sequenced to date (Nelson et al., 2008; Nelson, 2009). However, only a few members of the CYP81 family have been functionally characterized, linking them to specialized metabolism in response to biotic and abiotic stress (Akashi et al., 1998; Cabello-Hurtado et al., 1998; Pfalz et al., 2009; Kai et al., 2011).

For example, CYP81E members in *Medicago truncatula* catalyze regiospecific hydroxylations on ring B of the isoflavone skeleton in either the 2′ or 3′ positions, yielding products that are involved in pathogen defense and insect-induced responses, respectively (**Figure V-6a**) (Liu et al., 2003).

In *Arabidopsis thaliana*, members of the CYP81F subfamily are essential for the production of indole-type glucosinolates which accumulate upon pathogen attack (Pfalz et al., 2009; Pfalz et al., 2011). CYP81F2/CYP81F3 catalyze the C-hydroxylation of indol-3-yl-methyl glucosinolate (I3M) at position 4 to yield 4-hydroxy-indol-3-yl-methyl glucosinolate (4OH-I3M), while CYP81F4 catalyzed the *N*-hydroxylation of the same substrate to yield 1-hydroxy-indol-3-yl-methyl glucosinolate (1OH-I3M). Methylation of the introduced hydroxyl groups by indole glucosinolate methyltransferases (IGMT1 and IGMT2) gives rise to the final products 4MO-I3M and 1MO-I3M (**Figure V-6b**).

CYP81Q members catalyze stepwise formation of two methylenedioxy bridges in the biosynthesis of the lignan (+)-sesamin, which widely occurs in vascular plants (**Figure V-6c**) (Ono et al., 2006).

The previous examples show that different members of individual CYP81 subfamilies catalyze various reactions using the same substrate but show alternative regioselectivities yielding different products that are essential for plant defense.

Comparably, CYP81AAs utilize one substrate (2,4,6-triHB) and convert it into either 1,3,7- or 1,3,5-triHX. The ability of CYP81s to catalyze *N*-hydroxylation, methylenedioxy bridge formation and phenol coupling may suggest that other members of the same family are involved in catalyzing further unusual reactions yet to be characterized and encourage deeper investigation of the members of this family in different plants.

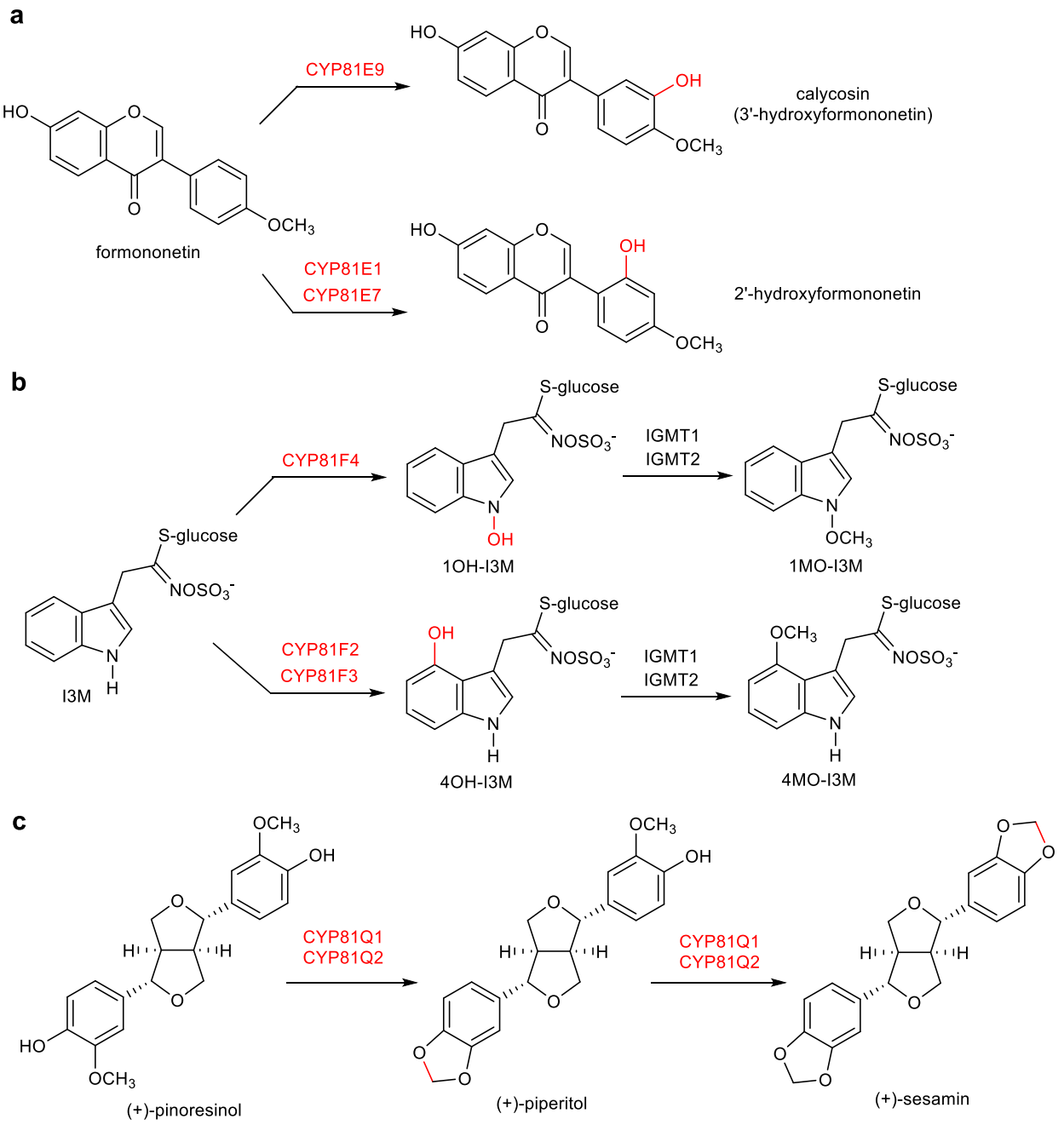


Figure V-6. Examples of reactions catalyzed by CYP81 family members. (**a**) Regiospecific hydroxylation of isoflavonoids by CYP81E subfamily members. (**b**) Regiospecific hydroxylation of indol-3-yl-methyl glucosinolate (I3M) by CYP81F subfamily members in the biosynthesis of 4-methoxy-indol-3-yl-methyl (4MO-I3M) and 1-methoxy-indol-3-yl-methyl (1MO-I3M) glucosinolates. (**c**) Consecutive formation of two methylenedioxy bridges during the biosynthesis of the lignan (+)-sesamin.

A search for sequences homologous to the isolated *CYP81AA*s in the transcriptomes of other xanthone-producing plants revealed the existence of sequences sharing more than 55 % amino acid identity with both CYP81AA1 and CYP81AA2, which therefore fall within the same subfamily

and can be classified as putative TXSs. These sequences were identified in families closely related to the Hypericaceae, namely the Clusiaceae and the Calophyllaceae. In contrast, transcriptomes of Gentianaceae species, which also produce xanthones, lacked *CYP81AA* homologs, indicating that other CYPs may be involved in xanthone biosynthesis in this family. Phylogenetic analysis of the sequences together with the isolated TXSs and representative members of other CYP81 subfamilies shows that all putative TXSs cluster together with the identified CYP81AAs in one clade with a bootstrap value of 100 (**Figure V-7**), indicating that they have evolved from a common ancestor. All other functionally related members in lignan, isoflavonoid or glucosinolate biosynthesis also clustered together in distinct clades with bootstrap values equal to or approaching 100 (**Figure V-7**).

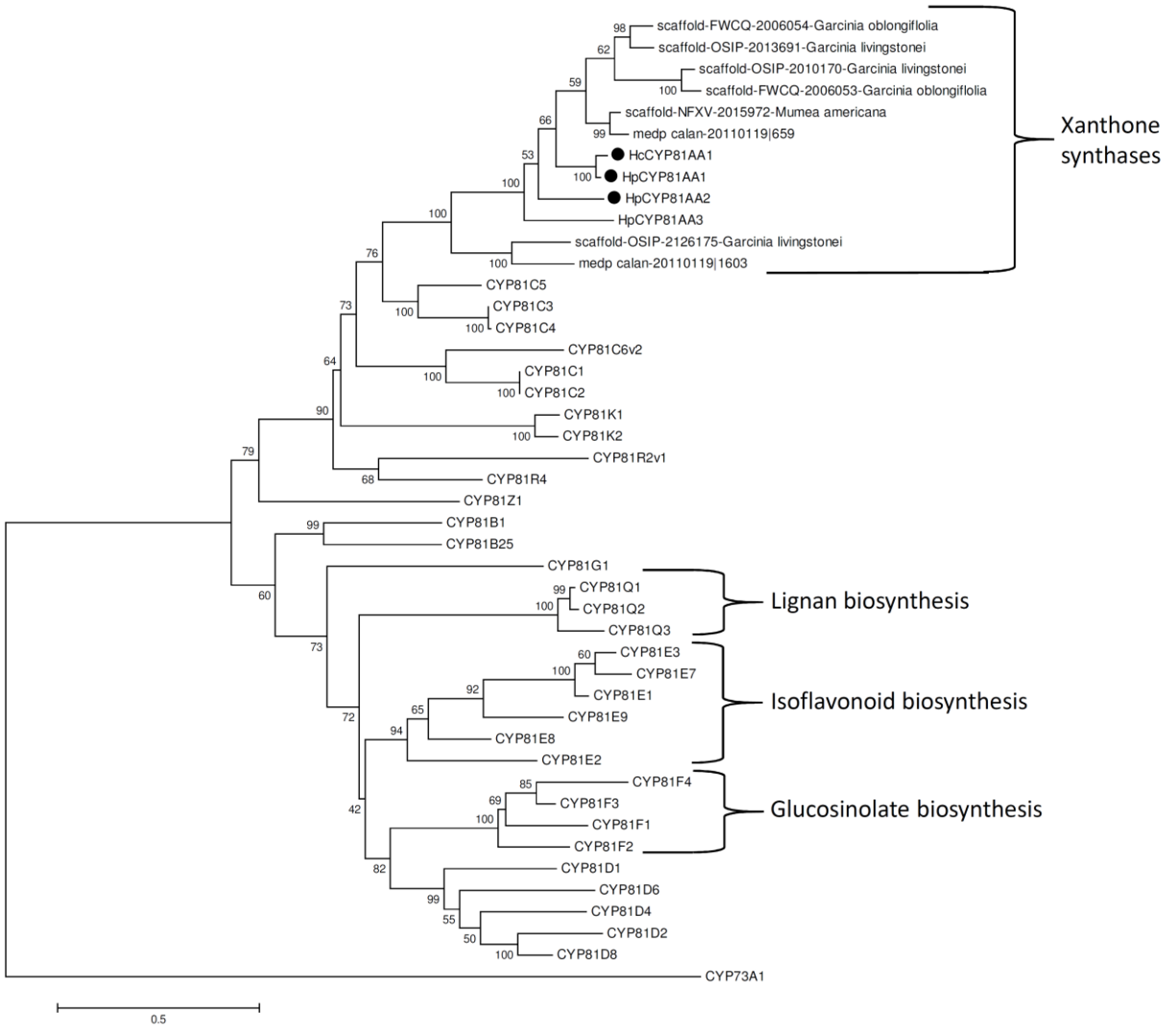


Figure V-7. Phylogenetic analysis of established and putative xanthone synthases together with representative members of the CYP81 family based on amino acid sequences. Tree reconstruction was performed with the maximum likelihood algorithm using MEGA 6 program. Bootstrap values (percent of 1000 replicates) for each cluster are shown at the nodes. CYP73A1 served as an outgroup to root the tree. The functionally characterized TXSs are designated by black circles.

3. Gene expression analysis

In *H. calycinum* cell cultures, expression of *HcCYP81AA1* (1,3,7-TXS) was induced by the addition of yeast extract as an elicitor and the increase in the transcript level preceded the accumulation of hyperxanthone E. Coordinated upregulation was not only observed for expression of *HcCYP81AA1* and *HcBPS*, which are directly associated to xanthone formation, but also for expression of *CNL*. Cinnamate:CoA ligase (CNL) catalyzes the upstream activation of cinnamic acid, thereby channeling the carbon flow from general phenylpropanoid metabolism to biosynthesis of the benzoyl unit, one of the building blocks of the xanthone scaffold (Gaid et al., 2012). In addition, the transcript level for *HcCPR2*, which indirectly contributes to hyperxanthone E biosynthesis, underwent similar changes.

Analysis of publicly available *H. perforatum* transcriptomes revealed that mainly roots and leaves express *CYP81AA1*, *CYP81AA2* and *BPS*. Roots of *Hypericum* species are the major site of xanthone accumulation while lower quantities are found in leaves (Avato, 2005).

4. Proposed mechanism for CYP-catalyzed phenol coupling reactions

The exact mechanism of CYP-catalyzed phenol couplings is not yet fully understood and cannot be easily rationalized in context of the traditional catalytic cycle of CYPs (Guengerich, 2007; Grobe et al., 2009). Hydroxylation by the oxygen rebound mechanism described in (I.1.4) is the classical CYP-catalyzed reaction, inserting one atom of molecular oxygen into the substrate and reducing the second to water (Munro et al., 2013). The cyclization step, however, involves an oxidative phenol coupling reaction, reducing both atoms of molecular oxygen to water (Mizutani and Sato, 2011). Two mechanisms have been outlined for the previously studied phenol coupling reactions.

4.1. The radical-pairing mechanism

The postulation of the radical-pairing mechanism for the formation of phenolic C–C or C–O–C bonds finds its origin back to the 1950s when Braton and Cohen (1957) suggested that the coupling proceeds via one-electron oxidation of phenols to generate a phenoxy radical. This radical possesses significant electron density at the *ortho* and *para* carbon atoms of the hydroxyl group as well as at the oxygen atom. Two radicals can afterwards undergo either intra- or intermolecular coupling to form new C–O or C–C bonds. New bonds to carbon atoms are formed exclusively at the *ortho* and *para* positions. Consistently, in natural products known to involve phenol coupling reactions between two aromatic rings during their biosynthesis, the reaction requires a free phenolic group to be located at the appropriate position in both precursor rings and occurs exclusively *ortho* and/or *para* to the respective hydroxyl groups (McDonald, 2008).

The intermolecular C–O phenol coupling between two molecules of (*S*) and (*R*)-*N*-methylcoclaurine catalyzed by berbamunine synthase (CYP80A1) is proposed to follow this mechanism (**Figure V-8a**). Two radicals are postulated to be generated by the Cpd I during two CYP oxidation cycles and coupling of the two species leads to the formation of berbamunine (Stadler and Zenk, 1993; Kraus and Kutchan, 1995). Likewise, the intramolecular C–C phenol coupling reaction of (*S*)-reticuline to corytuberine catalyzed by CYP80G2 in the plant *Coptis japonica* is postulated to follow the same mechanism (**Figure V-8b**). However, the generation of

the biradical species is proposed to occur within one catalytic cycle by Cpd I and Cpd II, respectively (Ikezawa et al., 2008).

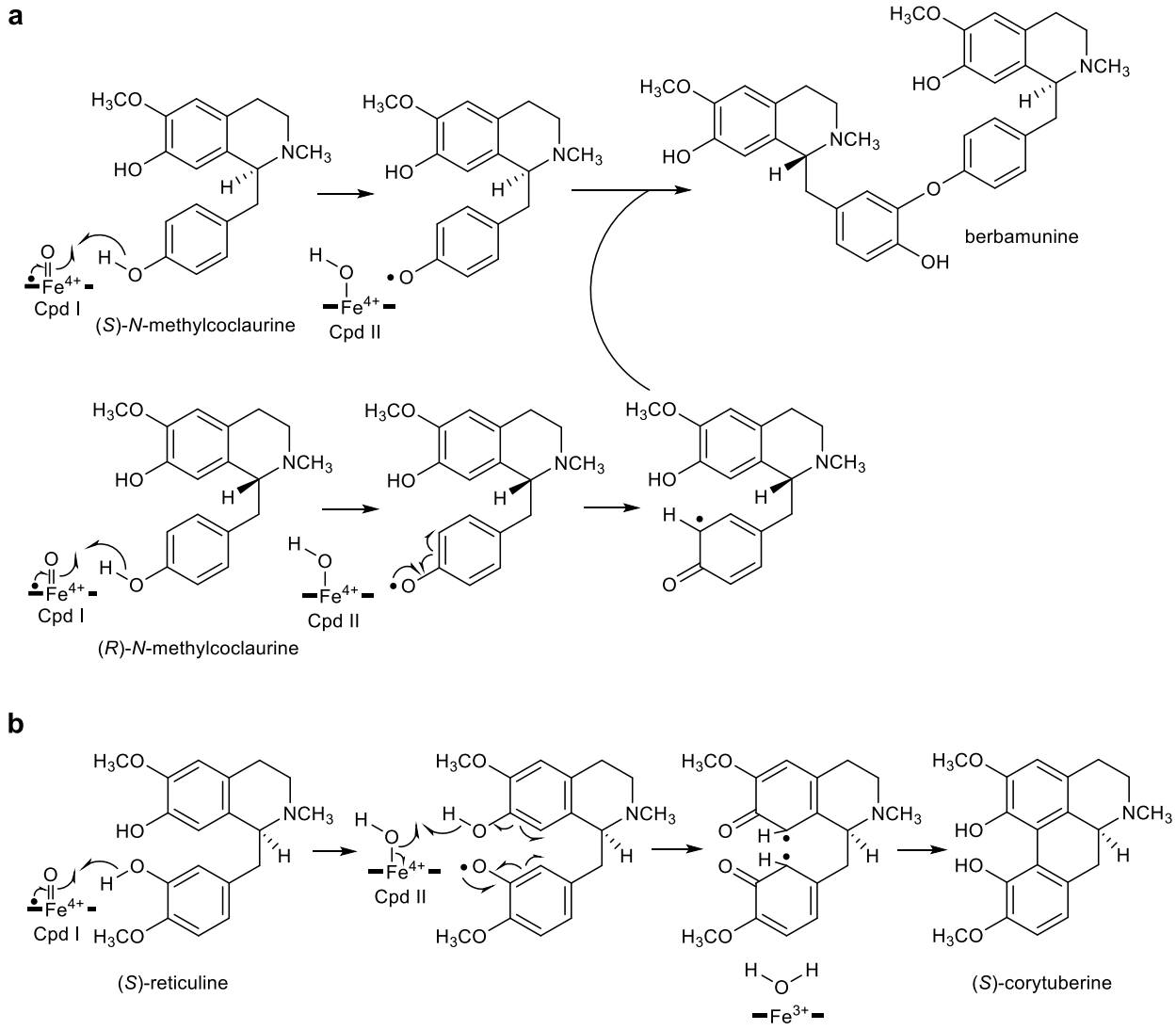


Figure V-8. (**a**) Postulated catalytic mechanism of oxidative phenol coupling of chiral benzyltetrahydroisoquinoline alkaloids involving two one-electron oxidation steps. Adapted from Stadler and Zenk (1993). (**b**) Proposed phenol coupling mechanism for the reaction catalyzed by CYP80G2 in *Coptis japonica*. The formation of (S)-corytuberine from (*S*)-reticuline is suggested to proceed by two one-electron transfers with subsequent diradical coupling within one CYP catalytic cycle; adapted from Ikezawa et al. (2008).

Recently, the C–C phenol coupling reaction catalyzed by CYP121 in *M. tuberculosis* was also proposed to proceed via the formation of a radical species on each tyrosyl moiety of cyclodityrosine (cYY), followed by radical coupling to form the cyclic dipeptide product mycocyclosin (**Figure V-9**) (Belin et al., 2009).

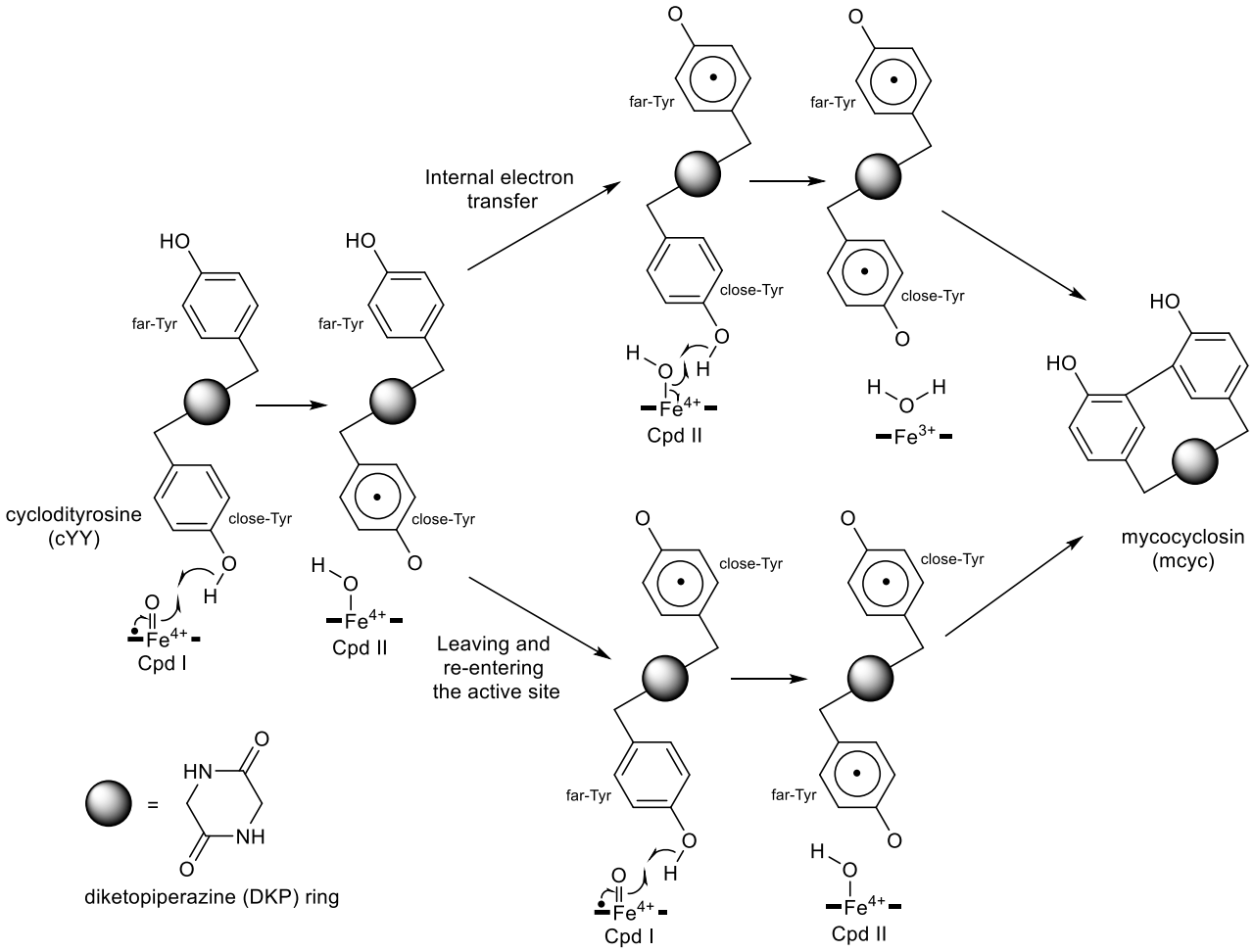


Figure V-9. Proposed phenol coupling mechanism for the formation of mycocyclosin from cyclodityrosine by CYP121 in *M. tuberculosis*; adapted from Belin et al. (2009).

The crystal structure of CYP121 bound to the substrate has been solved (Belin et al., 2009) and the mechanism was further investigated by quantum and molecular mechanics (Dumas et al., 2014). The crystal structure shows that one of the tyrosine moieties is located close to the heme (close-Tyr), while the other tyrosine is located far from it (far-Tyr). The close-Tyr is proposed to be initially oxidized by Cpd I to form the first radical. Several theories can explain the formation of the second radical on the far-Tyr. It is assumed that the mono radical may leave the active site and re-enter with the second Tyr (far-Tyr) facing the heme to yield the diradical species by the action of Cpd II from the same catalytic cycle or Cpd I from a new cycle. Another possibility is the internal electron transfer from the (close-Tyr) to the (far-Tyr) leading to the formation of a radical on the (far-Tyr) and then a new radical is formed on the (close-Tyr) by the action of Cpd I or Cpd II. The final step of radical coupling can occur inside or outside the protein core to yield the product (**Figure V-9**) (Dumas et al., 2014).

4.2. Single iron oxidation cycle mechanism

The formal diradical species suggested in the previous mechanism is energetically unstable. Therefore, the C–C coupling of (*R*)-reticuline to form salutaridine was favorably suggested to proceed within one CYP catalytic cycle, where the abstraction of a hydrogen by Cpd I leads to the formation of a phenoxy radical. Partial rotation of the substrate within the active site occurs, followed by a concomitant abstraction of a second hydrogen on the second aromatic ring by Cpd II together with electron transfer leading to the formation of the C–C bond. A water molecule is eliminated from the heme iron which returns to the resting state (**Figure V-10**). Finally, enolization of the intermediate yields the salutaridine product (Gesell et al., 2009; Grobe et al., 2009).

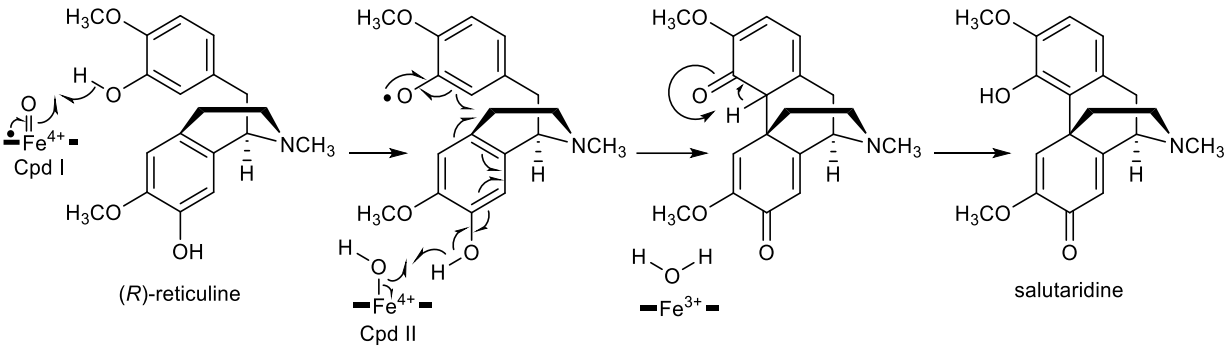


Figure V-10. Proposed phenol coupling mechanism for the reaction catalyzed by CYP719B1 in *Papaver somniferum* as well as CYP2D4 and CYP3A4 in mammals. The formation of salutaridine from (*R*)-reticuline is suggested to pass through a single cycle of iron oxidation; adapted from Grobe et al. (2009) and Gesell et al. (2009).

5. Proposed mechanisms for the reactions catalyzed by TXSs

As previously discussed, TXSs catalyze two consecutive reactions during xanthone biosynthesis. The first reaction is a conventional hydroxylation, while the second is a regioselective oxidative phenol coupling. Thus, TXSs can function as both monooxygenases and oxidases. The detection of 2,3′,4,6-tetraHB intermediate after incubation of either enzymes with 2,4,6-triHB suggests that this intermediate leaves the active site and re-enters in a conformation favorable for the cyclization reaction. In the first (monooxygenation) reaction, the rules of the classical catalytic cycle for CYPs apply. Cpd I abstracts a hydrogen from the non-hydroxylated aromatic ring to form a phenyl radical. Next, the formed Cpd II introduces a hydroxyl group at the 3′-position by the oxygen rebound mechanism (**Figure V-11a**). On the other hand, several theories may explain the second reaction in light of the previously studied phenol coupling mechanisms. Peters et al. (1997) proposed that the underlying reaction mechanism involves two one-electron oxidation steps. The first one-electron transfer and a deprotonation generate a phenoxy radical, whose electrophilic attack at C-2′ or C-6′ leads to the cyclization of the benzophenone. The intermediate hydroxycyclohexadienyl radical cation is transformed by loss of another electron and proton to 1,3,5- and 1,3,7-trihydroxyxanthones. However, the discovered ability of those enzymes to hydroxylate 2,4,6-triHB at the 3′ position and the fact that they introduce two hydroxyl groups on the non-hydroxylated ring of 2,4-diHB may afford revision of the previously hypothesized

mechanism. In my view, abstraction of a hydrogen atom from the monohydroxylated ring by Cpd I generates a resonance-stabilized radical with high electron density at C-2′/C-6′. Subsequently, instead of the classical oxygen rebound mechanism leading to the introduction of a hydroxyl group, a C–O bond is formed when this radical attacks the hydroxyl group at the C-2 position of the trihydroxylated ring and this hydroxyl group in concert undergoes abstraction of a hydrogen by Cpd II. This leads to the formation of the pyrone ring characteristic for the xanthone skeleton (**Figure V-11b**). Another possibility involves the radical-paring mechanism, where Cpd I and Cpd II form two radicals, one at the interchangeable 2/6-hydroxyl group and another at 2'/6'-position during the catalytic cycle. Radical coupling leads to cyclization of the benzophenone to a xanthone. However, as previously discussed by Grobe et al. (2009), the proposed formal diradical species would not be energetically stable. An alternative mechanism which was suggested by Kitanov and Nedialkov (2001) for the biosynthesis of xanthones involving water elimination from *ortho-ortho′*-dihydroxylated benzophenones might also be considered. However, the proposed intermediate was not detected in any of the TXSs-catalyzed reactions.

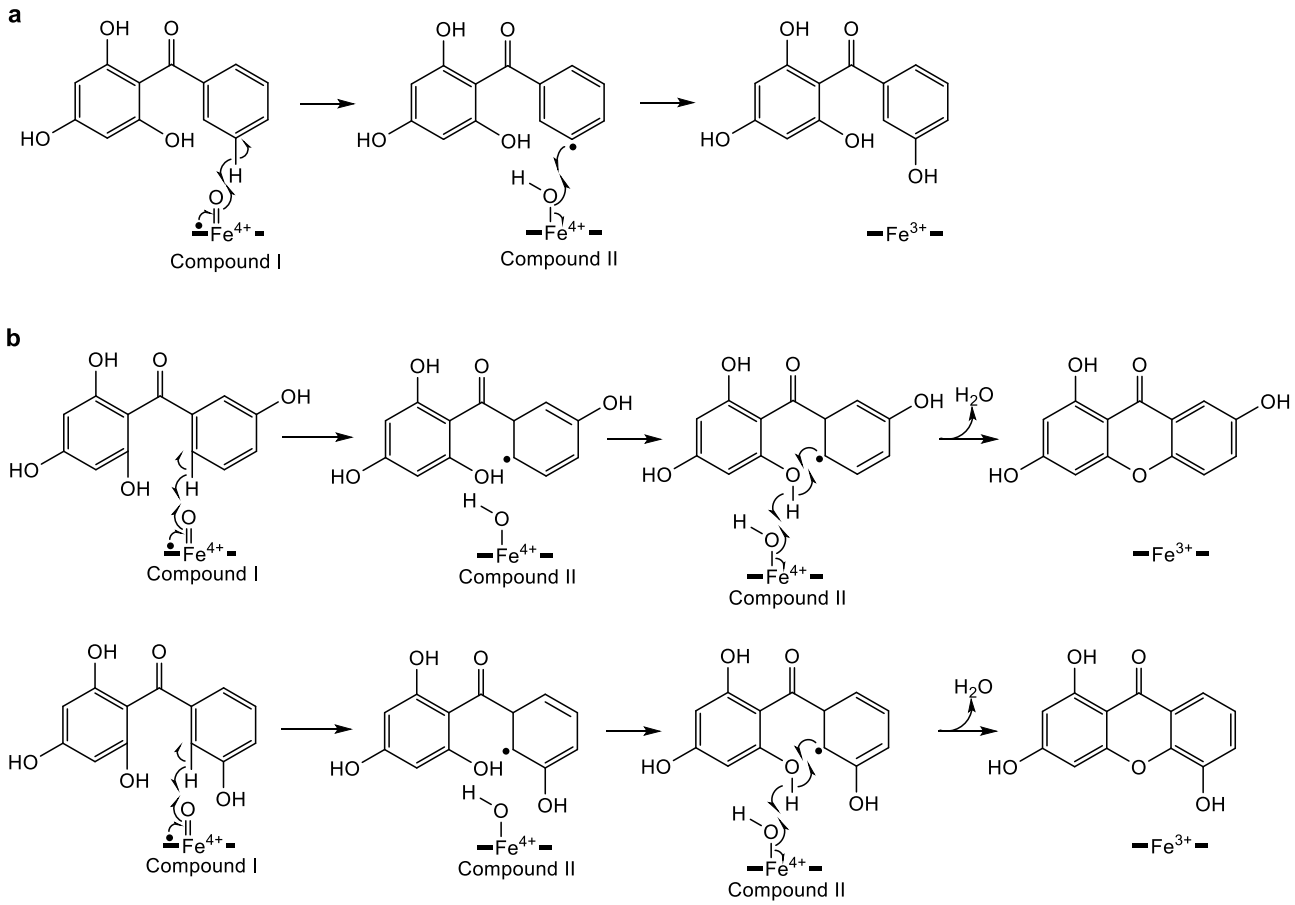


Figure V-11. Proposed reaction mechanisms of CYP81AA1 and CYP81AA2. (**a**) Hydroxylation reaction at the 3′-position catalysed by both enzymes. (**b**) Regioselective C–O phenol coupling reactions occurring either *para* or *ortho* to the introduced 3′-hydroxy group catalysed by CYP81AA1 and CYP81AA2, respectively.

Crystallization of each of the enzymes with the physiological substrates 2,4,6-triHB and 2,3′,4,6-tetraHB bound should reveal more about the decisive factors and the mechanism of CYP-catalyzed phenol coupling and answer the question why those enzymes can act as conventional hydroxylases with 2,4,6-triHB while performing C–O phenol couplings on the structurally related 2,3′,4,6-tetraHB.

6. NADPH-dependent cytochrome P450 reductase (CPR)

Most eukaryotic CYPs obtain their reducing equivalents from NADPH through CPR (Hannemann et al., 2007). In contrast to animals having a single CPR (Porter et al., 1990), each plant species has two or more CPRs that can be phylogenetically classified within two distinct classes (class I or class II) (Jensen and Moller, 2010). They differ mainly in their N-terminal membrane anchor sequences suggesting that they are forming metabolons together with different sets of CYPs localized on the membranes of endoplasmic reticulum (Jensen and Moller, 2010). Phylogenetic analysis of the identified *H. calycinum* CPR sequence together with 19 other CPR homologs from various plant species indicated that the identified *H. calycinum* CPR sequence falls within class II (**Figure V-12**).

It is worth mentioning that members of CPR class II are more abundant than those of class I. For example, all three rice CPRs, two of the three *Populus trichocarpa* (poplar) reductases (CPR2 and CPR3) as well as one of the two CPRs in *Gossypium hirsutum* (cotton) and *Arabidopsis thaliana* fall within class II CPRs (Schuler and Rupasinghe, 2011). In Arabidopsis, cotton and centaury, expression of class II *CPR* is induced by wounding, light and elicitation, whereas class I *CPR* is constitutively regulated (Mizutani and Ohta, 1998; Schwarz et al., 2009; Yang et al., 2010). Consistently, class II *CPR* transcripts were cloned from the SSH cDNA library of *H. calycinum* and isolated from cell cultures after elicitor treatment. Despite its indirect role in xanthone biosynthesis as an electron carrier between NADPH and TXSs, real-time qRT-PCR data demonstrated that the expression of HcCPR2 exhibited the same pattern as the other genes involved in xanthone biosynthesis in yeast-extract-treated *H. calycinum* cell cultures.

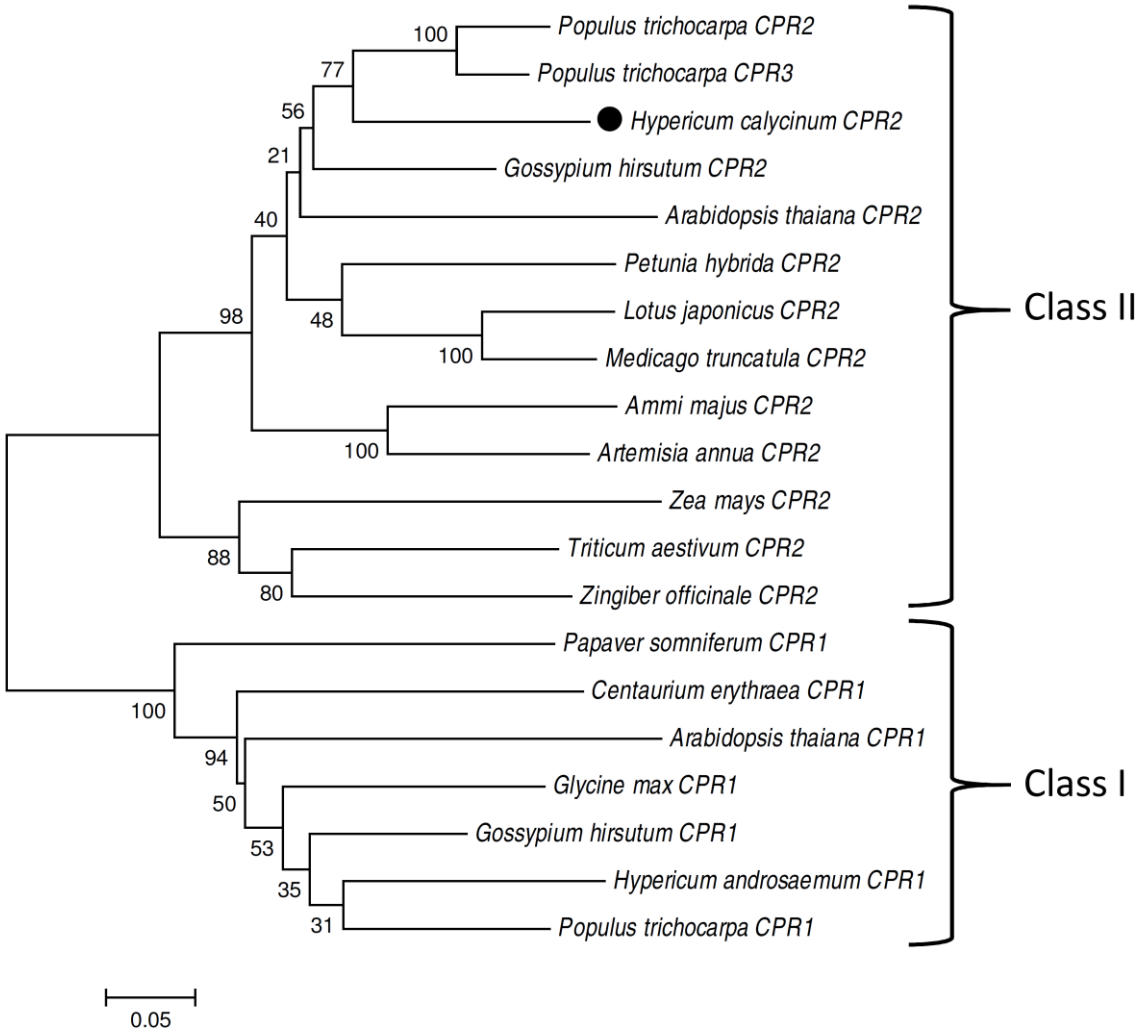


Figure V-12. Phylogenetic analysis of established and putative plant CPRs based on their amino acid sequences. Tree reconstruction was performed with the maximum likelihood algorithm using the MEGA 6 program. Bootstrap values (percent of 1000 replicates) for each cluster are shown at the nodes. *Ammi majus* CPR2 (AAS90127.1), *Arabidopsis thaliana* CPR1 (NP_194183.1), *A. thaliana* CPR2 (NP_849472.2), *Artemisia annua* CPR2 (ABI98819.1), *Centaurium erythraea* CPR1 (AAS92623.1), *Glycine max* CPR1 (NP_001236742.1), *Gossypium hirsutum* CPR1 (ACN54323.1), *G. hirsutum* CPR2 (ACN54324.1), *Hypericum androsaemum* CPR1 (AAS00459.1), *Lotus japonicus* CPR2 (BAG68945.1), *Medicago truncatula* CPR2 (XP_003610109.1), *Papaver somniferum* CPR1 (AAC05021.1), *Petunia hybrida* CPR2 (AAZ39649.1), *Populus trichocarpa* CPR1 (XP_002307336.1), *P. trichocarpa* CPR2 (AAK15260.1), *P. trichocarpa* CPR3 (AAK15261.1), *Triticum aestivum* CPR2 (CAC83301.1), *Zea mays* CPR2 (ACG28940.1), *Zingiber officinale* CPR2 (BAJ41269.1), *Homo sapiens* (NP_000932.3).

7. Identification of regioselectivity determinants

In spite of the low primary sequence identities between members of different CYP families that can reach as low as 20 %, they show distinct secondary structural elements and share a common fold. This allows modeling of CYPs using the available crystal structures even when they share relatively low identity on the primary sequence level. Most of the highly conserved structural elements of CYPs are involved in binding the heme prosthetic group and in activating molecular oxygen. However, due to the broad spectrum of substrates oxidized by CYPs with different

regio- and stereoselectivities, high sequence variability is observed in the regions responsible for interaction with substrates. Gotoh (1992) identified six substrate recognition sites (SRSs) in the crystal structure of CYP101A1 (P450cam) in complex with its substrate. Mutations targeting residues within those sites in various CYPs were shown to affect their substrate and product specificities.

Recently, a standard numbering system based on structure-guided sequence alignment was introduced for CYPs facilitating identification and comparisons between structurally corresponding residues in different proteins from the primary sequence data (Gricman et al., 2014). Identification of genes encoding TXSs with alternative regioselectivities offered the opportunity to explore their selectivity determinants. Sequence comparison, homology modeling and reciprocal site-directed mutagenesis demonstrated the impact of individual SRS residues on the selection of the alternative substrate conformers used. CYP81AA2 (1,3,5-TXS) was almost completely converted by mutagenesis to CYP81AA1 (1,3,7-TXS). Three amino acids in close proximity to the heme group (S375, L378 and A483; equivalent to the standard positions 328, 330.1 and 437, respectively; Gricman et al., 2014) were found to be most influential in orienting the coupling reaction *ortho* to the 3′-hydroxy group. S375 and L378 are located within SRS-5, which spans the loop between the EXXR motif and the *β*1-4 strand, whereas A483 belongs to SRS-6 (**Figure V-13, Figure V-14**). ;

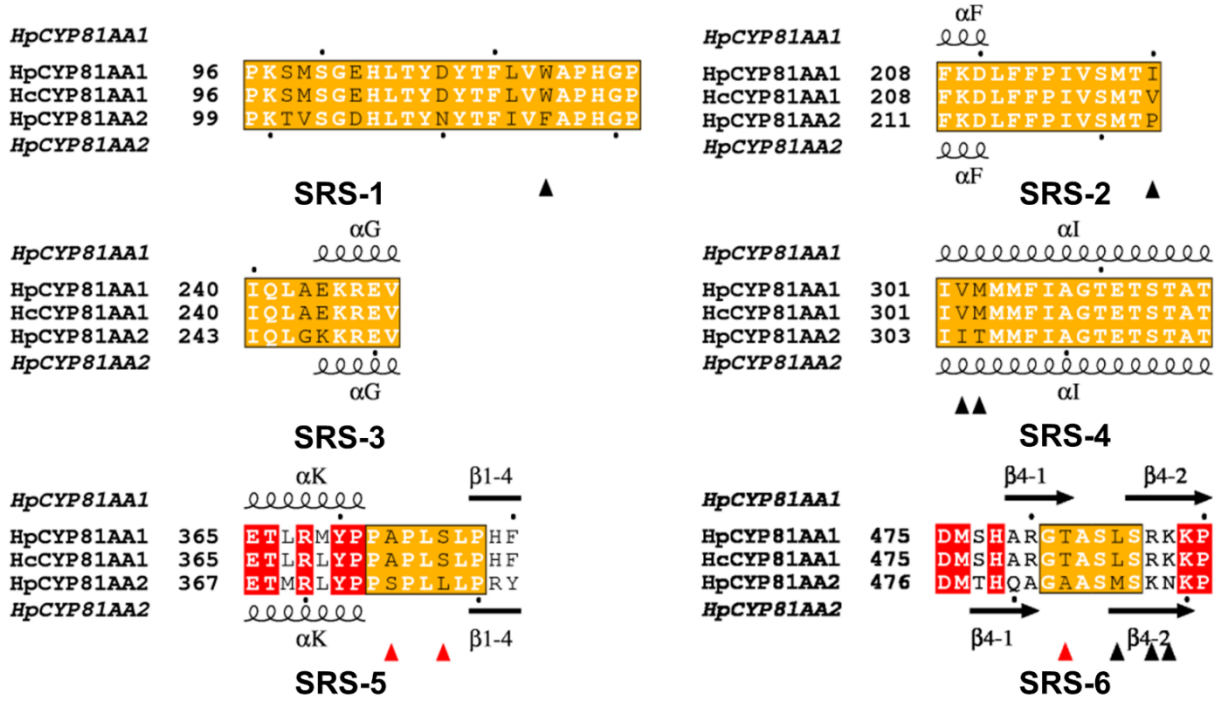


Figure V-13. Alignment of the SRS sequences of HpCYP81AA1, HcCYP81AA1, and HpCYP81AA2. Amino acids expected to be present in SRSs 1-6 are highlighted by orange boxes. Dissimilar residues are written in black letters and conserved residues outside the SRSs are represented in white with a red background. The conserved secondary structures of HpCYP81AA1 and HpCYP81AA2 are indicated at the top and the bottom, respectively. The black triangles point to positions selected for mutation and the red triangles refer to the major regioselectivity-determining residues in HpCYP81AA2.

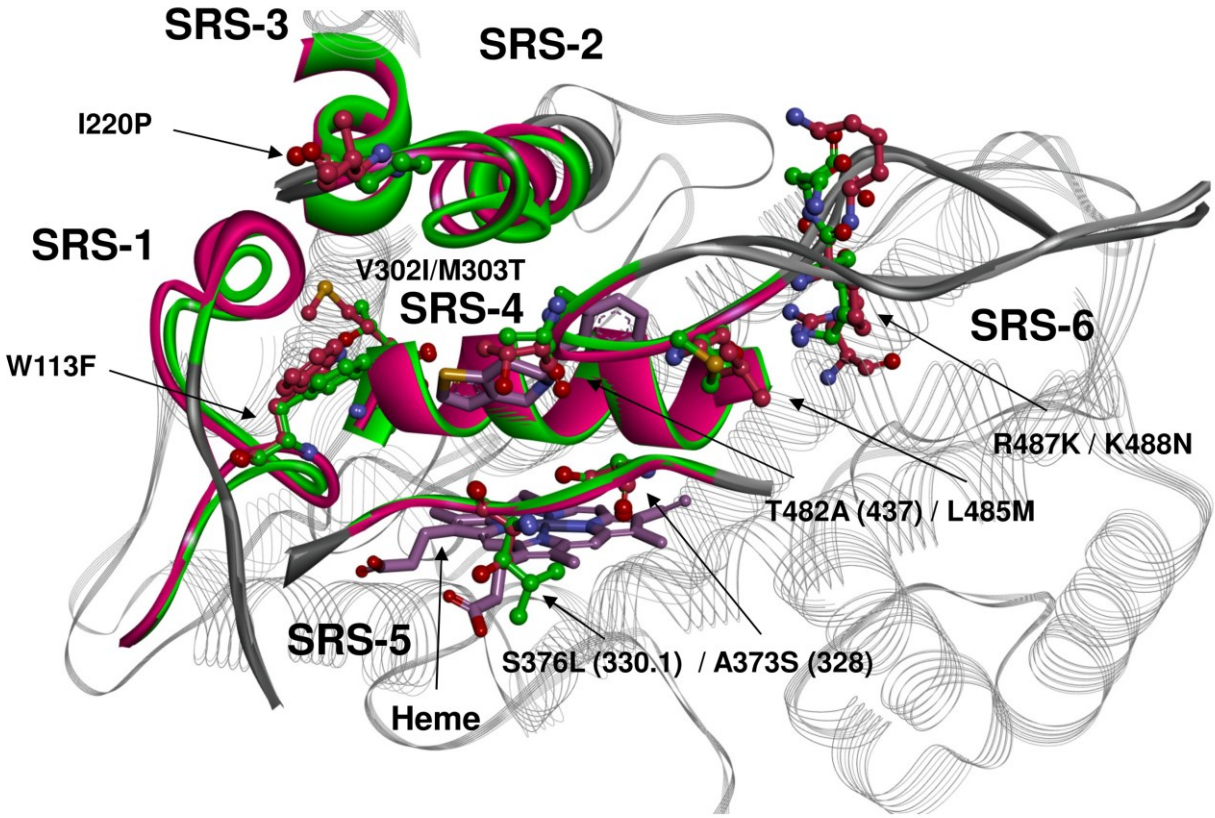


Figure V-14. Structural overlay of the homology models of HpCYP81AA1 (magenta, dark grey) and HpCYP81AA2 (green, light grey). Colored parts indicate the location of the six SRSs around the heme-containing active center with bound inhibitor ticlopidine (shown in stick representation) from the template X-ray structure 3kw4 (mammalian CYP2B4). Divergent residues are shown in ball and stick, labeling being based on CYP81AA1 numbering. Standard positions for major regioselectivity-determining residues are put in brackets.

Serine 375 (standard position 328) occupies position 5 behind the EXXR motif, its side chain pointing toward the heme group in the majority of CYPs structurally studied. The position is a mutation hotspot (Seifert and Pleiss, 2009; Gricman et al., 2015) and is preferentially occupied by a hydrophobic amino acid (Seifert and Pleiss, 2009). In CYP81AA2, however, there is a polar serine residue, which may be required for interaction with the polyhydroxylated substrates. In a number of CYPs, mutations at this standard position 328 resulted in dramatic changes in the regio- and stereoselectivities (Born et al., 1995; Schalk and Croteau, 2000; Kerdpin et al., 2004; Lentz et al., 2006). The point mutation F363I completely switched the regiospecificity of spearmint limonene 6-hydroxylase to that of peppermint limonene 3-hydroxylase (Schalk and Croteau, 2000). In CYP81AA2, however, the point mutation S375A caused only a partial change (~20 %) toward 1,3,7-triHX formation.

In addition to standard position 328, a second protruding residue in SRS-5 is expected to occupy either position 8 or 9 after the EXXR motif (corresponding to standard positions 331 or 332, respectively) (Seifert and Pleiss, 2009). Although L378 residue occupies position 8 behind the

EXXR motif, it was assigned the standard position 330.1 for which surprisingly no mutations have been reported. In contrast, multiple substitutions were carried out at the neighboring positions 330 and 331 (Gricman et al., 2015). It may contribute to the orientation of the substrate within the active site and/or shift the position of the whole loop by interaction with the heme group, affecting the size of the active site cavity.

SRS-6 is formed by the turn in *β*-sheet 4 between *β*4-1 and *β*4-2 toward the C-terminus. A483 (standard position 437) is located at the tip of this turn, its side chain protruding into the active site (**Figure V-14**). Mutations at this position affected the regiospecificities of multiple CYPs (Gricman et al., 2015), such as CYP94A2, in which the F494L substitution shifted the regioselectivity of fatty acid hydroxylation from ω to ω-1 (Kahn et al., 2001). In CYP81AA2, L378S and A483T exhibited less than 8% change each but induced around 20 and 29% shifts, respectively, in combination with S375A, indicating synergistic interaction in controlling the regioselectivity. The triple mutant S375A/L378S/A483T caused 80.7% of the total product to be released as 1,3,7-triHX and another 10.4% shift was contributed by the triple mutation M486L/K488R/N489K in SRS-6. However, no combination of mutations was capable of achieving the absolute regiospecificity of CYP81AA1. In case of the sextuple mutant S375A/L378S/A483T/M486L/K488R/N489K, a portion of 6.9% of the total product was still 1,3,5-triHX (beside 2% unidentified product). To summarize, S375A is the most influential mutation, however it cannot cause alone a complete shift in regioselectivity. There must be some "helper" mutations to achieve this. L378S and A483T are almost useless on their own or even in combination, but when added to S375A the effect is dramatic.

Reciprocal site-directed mutagenesis was also applied to CYP81AA1 (1,3,7-TXS). However, the maximum shift in the regioselectivity achieved was 6.8 %, indicating that subtle changes in the backbone rather than substitutions in the SRSs are needed to convert 1,3,7-TXS to 1,3,5-TXS. Similarly, the I364F modification in the peppermint limonene 3-hydroxylase, the reciprocal mutation to the above-mentioned substitution, did not achieve the regiospecificity of the spearmint limonene 6-hydroxylase, quite the contrary, it afforded an inactive, although properly folded, enzyme (Schalk and Croteau, 2000). Notably, none of the enzyme mutants generated exhibited uncoupling of the consecutive hydroxylation and coupling reactions.

CYPs are probably nature's most versatile enzymes in terms of substrate range and reaction type (Munro et al., 2013). This vast repertoire is here extended by the properties of CYP81AA1 and CYP81AA2, whose study provides further insight into CYP-catalyzed oxidative phenol coupling reactions. However, further work is needed to better understand the selectivity determinants, as indicated by the challenge to convert 1,3,7- to 1,3,5-TXS. As ring closure catalysts, the enzymes may be attractive for engineering efforts and biotechnological applications.

8. Perspectives

The identification of TXSs and studying their properties on the molecular level complete the cascade of enzymes required for the biosynthesis of xanthones in plants starting from malonyl- and benzoyl-CoA building blocks. This allows future engineering of the whole pathway in other hosts for the heterologous production of xanthones and their derivatives in microbes as well as for improving the desired values of other plants. Identification of the yet uncharacterized upstream steps for the biosynthesis of benzoyl-CoA from cinnamic acid would even allow the *de novo* biosynthesis of those products.

Another interesting aspect will be the identification of more TXS homologs from other xanthone-producing plants. Our phylogenetic analysis based on the currently available trascriptomic data revealed that members of the Malpighiales indeed express related genes. These genes should be isolated and functionally characterized to confirm the function of their protein products. On the other hand, no TXS homologs were identified in xanthone-producing plants from transcriptomes of the Gentianales. Although whole transcriptome data represent valuable sources of information, they cannot be relied on as the sole basis for the identification of new functional CYPs because of the large number of CYP superfamily members in plants and the absence of sequence to function correlations. Therefore, a strategy similar to that described in this work could be followed by using one or more members of the Gentianales to establish an analogous system to *H. calycinum* cell cultures that produces xanthones exclusively upon elicitation. Analysis of the gene expression data from this system together with the current knowledge gained in the present study about the composition of residues within the active site of TXSs would facilitate the identification of candidate genes encoding functional species-specific TXSs from the Gentianales.

Control of the hydroxylation pattern of the produced xanthones in the producing plants may be accomplished by manipulating the expression levels of the TXS enzymes. This might be beneficial in cases where one form, either 1,3,5- or 1,3,7-triHX derivatives, is desired because of its pharmacological activity. When both enzymes are equally expressed, they would compete for the benzophenone substrate and produce a mixture of both 1,3,7- and 1,3,5- derivatives, which would subsequently compete for the next enzyme in the pathway, producing less amounts of the desired product. Moreover, knocking out/down all TXSs in a xanthone-producing plant may increase the benzophenone pool available for the biosynthesis of bioactive and structurally intriguing prenylated benzophenone derivatives, such as sampsoniones (Hu and Sim, 1999; Hu and Sim, 2000).

The trihydroxyxanthones produced can be further hydroxylated, methoxylated and prenylated at different positions to produce more complex structures. The enzymes catalyzing those steps are interesting targets for future research (Schmidt et al., 2000; Fiesel et al., 2015). Studying whether or not those enzymes are associated in gene clusters would reveal valuable information about the evolution of this metabolic pathway and its regulation. It would also facilitate the identification of further genes involved in upstream and downstream steps of the pathway. However, this requires whole genome sequencing data with annotation of the genes to the different chromosomes, which is to date not available for xanthone-producing species.

It has been reported that many sequential enzymes involved in certain biosynthetic pathways are structured in groups, referred to as metabolons. They are clusters of proteins which are associated together through non-covalent interaction and often interact with cellular membranes. This structuring facilitates rapid and localized channeling of intermediates between consecutive enzymes in biosynthetic pathways. It has been proposed that in many cases CYPs act as nuclei on which the other proteins associate to establish metabolons (Ralston and Yu, 2006). Therefore, it is interesting to examine if the TXSs can act as cores, upon which other enzymes of the xanthone biosynthetic pathway can form metabolons.

Although homology modeling was useful in the identification of regioselectivity determinants in CYP81AA2, manipulating the regiospecificity of CYP81AA1 remains challenging. Crystallization of the membrane anchor-freed TXSs is necessary to gain deeper insight into the structural elements responsible for the regioselectivity as well as the reaction mechanisms underlying the observed bifunctionality.

VI. Summary

- Xanthones constitute a group of natural products found in fungi, lichens and some higher plants. They may contribute, among other constituents such as acylphloroglucinol derivatives, to the overall effects of the medicinal plant *H. perforatum*.
- Extracts of common St. John's wort (*H. perforatum*) are among the top selling herbal products in the USA and Europe and are widely used in the treatment of mild to moderate depression.
- In a number of other plants, the xanthone scaffold is the core of many complex hydroxylated, methoxylated, glycosylated and prenylated polycyclic compounds. Xanthones possess a wide array of pharmacological activities including potential anti-Alzheimer properties (Zhang, 2005; Sethiya and Mishra, 2014).
- A crucial branching point in plant xanthone biosynthesis is the cyclization of 2,3′,4,6-tetraHB via regioselective intramolecular oxidative C–O phenol couplings either *para* or *ortho* to the 3′-hydroxyl group to yield 1,3,7- and 1,3,5-triHX, respectively (Peters et al., 1997).
- The enzymes involved in the cyclization reactions have been biochemically identified as cytochrome P450 enzymes (CYPs). However, no cDNAs encoding these enzymes have been cloned so far.
- CYPs constitute a superfamily of heme-thiolate enzymes which normally catalyze hydroxylation/monooxygenation reactions. However, some members catalyze less common unusual reactions, such as C–O phenol couplings.
- Yeast extract-treated *H. calycinum* cell cultures produce hyperxanthone E which is a derivative of 1,3,7-triHX, while control (non-treated) cells fail to produce this compound (Gaid et al., 2012). This system was utilized to construct a suppression subtractive hybridization cDNA library enriched in elicitor-induced transcripts.
- Bioinformatic analysis of the ESTs in the library followed by rapid amplification of cDNA ends (RACE) led to the isolation of full-length cDNAs encoding HcCYP81AA1 and HcCPR2.
- The microsomal fraction of yeast cells expressing *HcCYP81AA1* and *HcCPR2* catalyzed the following reactions:
  - Cyclization of 2,3′,4,6-tetraHB into 1,3,7-triHX,
  - Hydroxylation of 2,4,6-triHB at the 3′ position and subsequent cyclization of the resulting 2,3′,4,6-tetraHB into 1,3,7-triHX,
  - Hydroxylation of 2,4-diHB at the 3′ position, then further hydroxylation of the resulting 2,3′,4-triHB *para* to the introduced hydroxyl group to yield 2,2′,4,5′-tetraHB,
  - Cyclization of 2,3′,4,4′,6-pentaHB into 1,3,6,7-tetraHX.
- HcCYP81AA1 was identified as a bifunctional enzyme possessing both benzophenone 3′-hydroxylase and 1,3,7-trihydroxyxanthone synthase activities.
- Gene expression analysis by real-time qRT-PCR revealed coordinated expression of the *HcBPS*, *HcCYP81AA1* and *HcCPR2* genes. The expression profiles came in accordance with the published accumulation pattern of hyperxanthone E (Gaid et al., 2012).

- The sequence of HcCYP81AA1 was used to query the public *H. perforatum* transcriptome in the Medicinal Plant Genomics Resource (MPGR), leading to the identification of three homologous sequences, namely HpCYP81AA1, HpCYP81AA2 and HpCYP81AA3.
- Microsomes of yeast cells expressing *HpCYP81AA1* with *HcCPR2* showed an activity profile identical to that of HcCYP81AA1.
- The microsomal fraction of yeast cells expressing *HcCYP81AA2* and *HcCPR2* catalyzed the following reactions:
  - Cyclization of 2,3′,4,6-tetraHB into 1,3,5-triHX,
  - Hydroxylation of 2,4,6-triHB at the 3′ position and subsequent cyclization of the resulting 2,3′,4,6-tetraHB into 1,3,5-triHX,
  - Hydroxylation of 2,4-diHB at the 3′ position, then further hydroxylation of the resulting 2,3′,4-triHB *ortho* to the introduced hydroxyl group to yield 2,2′,3,4′-tetraHB,
  - Cyclization of 2,3′,4,4′,6-pentaHB into 1,3,6,7-tetraHX.
- HpCYP81AA2 was identified as a bifunctional enzyme possessing both benzophenone 3′-hydroxylase and 1,3,5-trihydroxyxanthone synthase activities.
- Microsomes of yeast cells expressing *HpCYP81AA3* and *HcCPR2* could only catalyze the cyclization of 2,3′,4,4′,6-pentaHB into 1,3,6,7-tetraHX.
- A mechanism was suggested to explain the hydroxylation and the alternative phenol coupling reaction steps.
- Phylogenetic analysis of putative TXS sequences from the transcriptomes of other xanthone-producing plants, the isolated CYP81AAs and representative members of CYP81 subfamilies showed that all putative TXSs cluster together with the identified CYP81AAs in one clade, indicating their evolution from a common ancestor.
- Reciprocal site-directed mutagenesis based on sequence comparisons, analysis of the substrate recognition sites (SRSs) and homology models built for both enzymes led to the identification of three residues (Ser 375, Leu 378 in SRS-5 and Ala 483 in SRS-6), which mainly control the regioselectivity of CYP81AA2 (1,3,5-TXS). This enzyme was almost completely converted to 1,3,7-TXS by a sextuple mutation.
- The regiospecificity of CYP81AA1 was not significantly affected by mutations performed on the corresponding residues or any other mutation attempted in the assigned SRSs, indicating the involvement of further elements outside the known CYP SRSs in controlling the regioselectivety of CYP81AA1.
- The efforts reported herein will enable better understanding of the selectivity determinants in CYP-catalyzed reactions, give a deeper insight into the mechanism of phenol couplings catalyzed by CYPs and allow for future generation of transgenic xanthone-producing plants with customized products. Moreover, the characterized CYPs have the potential to be utilized as ring closure catalysts in the biotechnological production of xanthones and xanthone-related natural and unnatural products.

VII. References

Abe F, Nagafuji S, Okabe H, Higo H, Akahane H (2003) Trypanocidal constituents in plants 2. Xanthones from the stem bark of *Garcinia subelliptica*. Biol Pharm Bull **26:** 1730-1733

Akashi T, Aoki T, Ayabe S (1998) CYP81E1, a cytochrome P450 cDNA of licorice (*Glycyrrhiza echinata* L.), encodes isoflavone 2'-hydroxylase. Biochem Biophys Res Commun **251:** 67-70

Akashi T, Aoki T, Ayabe S (1999) Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. Plant Physiol **121:** 821-828

Al-Massarani SM, El Gamal AA, Al-Musayeib NM, Mothana RA, Basudan OA, Al-Rehaily AJ, Farag M, Assaf MH, El Tahir KH, Maes L (2013) Phytochemical, antimicrobial and antiprotozoal evaluation of *Garcinia mangostana* pericarp and alpha-mangostin, its major xanthone derivative. Molecules **18:** 10599-10608

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389-3402

APGIII (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc **161:** 105-121

Avato P (2005) A survey on the *Hypericum* genus: Secondary metabolites and bioactivity. *In* R Atta ur, ed, Studies in Natural Products Chemistry, Vol Volume 30. Elsevier, pp 603-634

Bak S, Beisson F, Bishop G, Hamberger B, Hofer R, Paquette S, Werck-Reichhart D (2011) Cytochromes P450. Arabidopsis Book **9:** e0144

Becker DM, Lundblad V (2001) Introduction of DNA into yeast cells. Curr Protoc Mol Biol

Belin P, Le Du MH, Fielding A, Lequin O, Jacquet M, Charbonnier J-B, Lecoq A, Thai R, Courçon M, Masson C, Dugave C, Genet R, Pernodet J-L, Gondry M (2009) Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A **106:** 7426-7431

Bennett GJ, Lee H-H (1989) Xanthones from Guttiferae. Phytochemistry **28:** 967-998

Bilia AR, Gallori S, Vincieri FF (2002) St. John's wort and depression: Efficacy, safety and tolerability-an update. Life Sci **70:** 3077-3096

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res **7:** 1513-1523

Bischoff D, Pelzer S, Bister B, Nicholson GJ, Stockert S, Schirle M, Wohlleben W, Jung G, Süssmuth RD (2001) The Biosynthesis of vancomycin-type glycopeptide antibiotics—The order of the cyclization steps. Angew Chem Int Ed Engl **40:** 4688-4691

Born SL, John GH, Harlow GR, Halpert JR (1995) Characterization of the progesterone 21-hydroxylase activity of canine cytochrome P450 PBD-2/P450 2B11 through reconstitution, heterologous expression, and site-directed mutagenesis. Drug Metab Dispos **23:** 702-707

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:** 248-254

Braton DHR, Cohen T (1957) In: Festschrift Dr A Stoll, Birkhäuser, Basel

Butterweck V (2003) Mechanism of action of St John's wort in depression : what is known? CNS Drugs **17:** 539-562

Buttner M, Barleben L (2012) One-pot fusion polymerase chain reaction for combinatorial synthesis of DNA from several cassettes. Anal Biochem **421:** 797-798

Cabello-Hurtado F, Batard Y, Salaün J-P, Durst F, Pinot F, Werck-Reichhart D (1998) Cloning, expression in yeast, and functional characterization of CYP81B1, a plant cytochrome P450 that catalyzes in-chain hydroxylation of fatty acids. J Biol Chem **273:** 7260-7267

Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, Graham N, Odoardi M, Piano E, Arcioni S, May S, Scotti C, Calderini O (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell **23:** 3070-3081

Chen LG, Yang LL, Wang CC (2008) Anti-inflammatory activity of mangostins from *Garcinia mangostana*. Food Chem Toxicol **46:** 688-693

Chen SX, Wan M, Loh BN (1996) Active constituents against HIV-1 protease from *Garcinia mangostana*. Planta Med **62:** 381-382

Chericoni S, Testai L, Calderone V, Flamini G, Nieri P, Morelli I, Martinotti E (2003) The xanthones gentiacaulein and gentiakochianin are responsible for the vasodilator action of the roots of *Gentiana kochiana*. Planta Med **69:** 770-772

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21:** 3674-3676

Crockett SL, Poller B, Tabanca N, Pferschy-Wenzig E-M, Kunert O, Wedge DE, Bucar F (2011) Bioactive xanthones from the roots of *Hypericum perforatum* (common St John's wort). J Sci Food Agric **91:** 428-434

Crockett SL, Robson NK (2011) Taxonomy and chemotaxonomy of the genus *Hypericum*. Med Aromat Plant Sci Biotechnol **5:** 1-13

Cryle MJ, Staaden J, Schlichting I (2011) Structural characterization of CYP165D3, a cytochrome P450 involved in phenolic coupling in teicoplanin biosynthesis. Arch Biochem Biophys **507:** 163-173

Danielson PB (2002) The cytochrome P450 superfamily: Biochemistry, evolution and drug metabolism in humans. Curr Drug Metab **3:** 561-597

Deachathai S, Mahabusarakam W, Phongpaichit S, Taylor WC (2005) Phenolic compounds from the fruit of *Garcinia dulcis*. Phytochemistry **66:** 2368-2375

Decosterd LA, Hoffmann E, Kyburz R, Bray D, Hostettmann K (1991) A new phloroglucinol derivative from *Hypericum calycinum* with antifungal and *in vitro* antimalarial activity. Planta Med **57:** 548-551

Decosterd LA, Stoeckli-Evans H, Chapuis J-C, Sordat B, Hostettmann K (1989) New cell growth-inhibitory cyclohexadienone derivatives from *Hypericum calycinum* L. Helv Chim Acta **72:** 1833-1845

Demirkiran O (2007) Xanthones in *Hypericum*: Synthesis and biological activities. *In* MTH Khan, ed, Bioactive Heterocycles III, Vol 9. Springer Berlin Heidelberg, pp 139-178

Di Carlo G, Borrelli F, Ernst E, Izzo AA (2001) St John's wort: Prozac from the plant kingdom. Trends Pharmacol Sci **22:** 292-297

Dohmen RJ, Strasser AW, Honer CB, Hollenberg CP (1991) An efficient transformation procedure enabling long-term storage of competent cells of various yeast genera. Yeast **7:** 691-692

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem **24:** 1999-2012

Dumas VG, Defelipe LA, Petruk AA, Turjanski AG, Marti MA (2014) QM/MM study of the C–C coupling reaction mechanism of CYP121, an essential cytochrome P450 of *Mycobacterium tuberculosis*. Proteins **82:** 1004-1021

El-Seedi HR, El-Barbary MA, El-Ghorab DM, Bohlin L, Borg-Karlson AK, Goransson U, Verpoorte R (2010) Recent insights into the biosynthesis and biological activities of natural xanthones. Curr Med Chem **17:** 854-901

El-Seedi HR, El-Ghorab DM, El-Barbary MA, Zayed MF, Goransson U, Larsson S, Verpoorte R (2009) Naturally occurring xanthones; latest investigations: isolation, structure elucidation and chemosystematic significance. Curr Med Chem **16:** 2581-2626

Felth J, Lesiak-Mieczkowska K, D'Arcy P, Haglund C, Gullbo J, Larsson R, Linder S, Bohlin L, Fryknas M, Rickardson L (2013) Gambogic acid is cytotoxic to cancer cells through inhibition of the ubiquitin-proteasome system. Invest New Drugs **31:** 587-598

Field B, Osbourn AE (2008) Metabolic diversification—independent assembly of operon-like gene clusters in different plants. Science **320:** 543-547

Fiesel T, Gaid M, Muller A, Bartels J, El-Awaad I, Beuerle T, Ernst L, Behrends S, Beerhues L (2015) Molecular cloning and characterization of a xanthone prenyltransferase from *Hypericum calycinum* cell cultures. Molecules **20:** 15616-15630

Fotie J, Bohle DS (2006) Pharmacological and biological activities of xanthones. Antiinfect Agents Med Chem **5:** 15-31

Franklin G, Conceicao LF, Kombrink E, Dias AC (2009) Xanthone biosynthesis in *Hypericum perforatum* cells provides antioxidant and antimicrobial protection upon biotic stress. Phytochemistry **70:** 60-68

Fukai T, Oku Y, Hou AJ, Yonekawa M, Terada S (2005) Antimicrobial activity of isoprenoid-substituted xanthones from *Cudrania cochinchinensis* against vancomycin-resistant enterococci. Phytomedicine **12:** 510-513

Gaid MM, Sircar D, Muller A, Beuerle T, Liu B, Ernst L, Hansch R, Beerhues L (2012) Cinnamate:CoA ligase initiates the biosynthesis of a benzoate-derived xanthone phytoalexin in *Hypericum calycinum* cell cultures. Plant Physiol **160:** 1267-1280

Gay SC, Roberts AG, Maekawa K, Talakad JC, Hong WX, Zhang Q, Stout CD, Halpert JR (2010) Structures of cytochrome P450 2B4 complexed with the antiplatelet drugs ticlopidine and clopidogrel. Biochemistry **49:** 8709-8720

Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF (2006) Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biol **4:** 41

Gesell A, Rolf M, Ziegler J, Diaz Chavez ML, Huang FC, Kutchan TM (2009) CYP719B1 is salutaridine synthase, the C–C phenol-coupling enzyme of morphine biosynthesis in opium poppy. J Biol Chem **284:** 24432-24442

Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem **267:** 83-90

Graham SE, Peterson JA (1999) How similar are P450s and what can their differences teach us? Arch Biochem Biophys **369:** 24-29

Gricman L, Vogel C, Pleiss J (2014) Conservation analysis of class-specific positions in cytochrome P450 monooxygenases: Functional and structural relevance. Proteins **82:** 491-504

Gricman L, Vogel C, Pleiss J (2015) Identification of universal selectivity-determining positions in cytochrome P450 monooxygenases by systematic sequence-based literature mining. Proteins **83:** 1593-1603

Grobe N, Zhang B, Fisinger U, Kutchan TM, Zenk MH, Guengerich FP (2009) Mammalian cytochrome P450 enzymes catalyze the phenol-coupling step in endogenous morphine biosynthesis. J Biol Chem **284:** 24425-24431

Guengerich FP (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. Chem Res Toxicol **14:** 611-650

Guengerich FP (2007) Mechanisms of cytochrome P450 substrate oxidation: MiniReview. J Biochem Mol Toxicol **21:** 163-168

Guengerich FP, Martin MV, Sohl CD, Cheng Q (2009) Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. Nat Protoc **4:** 1245-1251

Guengerich FP, Munro AW (2013) Unusual cytochrome P450 enzymes and reactions. J Biol Chem **288:** 17065-17073

Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Trans R Soc Lond B Biol Sci **368:** 20120426

Hannemann F, Bichet A, Ewen KM, Bernhardt R (2007) Cytochrome P450 systems—biological variations of electron transport chains. Biochim Biophys Acta **1770:** 330-344

Haslinger K, Maximowitsch E, Brieke C, Koch A, Cryle MJ (2014) Cytochrome P450 $OxyB_{tei}$ catalyzes the first phenolic coupling step in teicoplanin biosynthesis. ChemBioChem **15:** 2719-2728

Haslinger K, Peschke M, Brieke C, Maximowitsch E, Cryle MJ (2015) X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. Nature **521:** 105-109

Hay AE, Helesbeux JJ, Duval O, Labaied M, Grellier P, Richomme P (2004) Antimalarial xanthones from *Calophyllum caledonicum* and *Garcinia vieillardii*. Life Sci **75:** 3077-3085

Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ (2001) The CYP88A cytochrome P450, *ent*-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. Proc Natl Acad Sci U S A **98:** 2065-2070

Henderson L, Yue QY, Bergquist C, Gerden B, Arlett P (2002) St John's wort (*Hypericum perforatum*): drug interactions and clinical outcomes. Br J Clin Pharmacol **54:** 349-356

Hoffman CS, Winston F (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene **57:** 267-272

Hu L-H, Sim K-Y (2000) Sampsoniones A–M, a unique family of caged polyprenylated benzoylphloroglucinol derivatives, from *Hypericum sampsonii*. Tetrahedron **56:** 1379-1386

Hu LH, Sim KY (1999) Cytotoxic polyprenylated benzoylphloroglucinol derivatives with an unusual adamantyl skeleton from *Hypericum sampsonii* (Guttiferae). Org Lett **1:** 879-882

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res **9:** 868-877

Iinuma M, Tosa H, Tanaka T, Asai F, Kobayashi Y, Shimano R, Miyauchi K (1996) Antibacterial activity of xanthones from guttiferaeous plants against methicillin-resistant *Staphylococcus aureus*. J Pharm Pharmacol **48:** 861-865

Ikezawa N, Iwasa K, Sato F (2007) Molecular cloning and characterization of methylenedioxy bridge-forming enzymes involved in stylopine biosynthesis in *Eschscholzia californica*. FEBS J **274:** 1019-1035

Ikezawa N, Iwasa K, Sato F (2008) Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C–C phenol coupling of (*S*)-reticuline in magnoflorine biosynthesis, from cultured *Coptis japonica* cells. J Biol Chem **283:** 8810-8821

Ikezawa N, Tanaka M, Nagayoshi M, Shinkyo R, Sakaki T, Inouye K, Sato F (2003) Molecular cloning and characterization of CYP719, a methylenedioxy bridge-forming enzyme that belongs to a novel P450 family, from cultured *Coptis japonica* cells. J Biol Chem **278:** 38557-38565

Irmler S, Schröder G, St-Pierre B, Crouch NP, Hotze M, Schmidt J, Strack D, Matern U, Schröder J (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. Plant J **24:** 797-804

Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem **23:** 1623-1641

Jefferson A, Scheinmann F (1965) Presence of 1,3,6,7-tetrahydroxyxanthone in maclurin from *Chlorophora tinctoria* (L) Gaud. (*Morus tinctoria* L) (Moraceae). Nature **207:** 1193-1193

Jensen K, Moller BL (2010) Plant NADPH-cytochrome P450 oxidoreductases. Phytochemistry **71:** 132-141

Jiang DJ, Dai Z, Li YJ (2004) Pharmacological effects of xanthones as cardiovascular protective agents. Cardiovasc Drug Rev **22:** 91-102

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol **292:** 195-202

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci **8:** 275-282

Kahn RA, Bouquin RL, Pinot F, Benveniste I, Durst F (2001) A conservative amino acid substitution alters the regiospecificity of CYP94A2, a fatty acid hydroxylase from the plant *Vicia sativa*. Arch Biochem Biophys **391:** 180-187

Kai K, Takahashi H, Saga H, Ogawa T, Kanaya S, Ohta D (2011) Metabolomic characterization of the possible involvement of a cytochrome P450, CYP81F4, in the biosynthesis of indolic glucosinolate in *Arabidopsis*. Plant Biotechnol **28:** 379-385

Kasper S, Caraci F, Forti B, Drago F, Aguglia E (2010) Efficacy and tolerability of *Hypericum* extract for the treatment of mild to moderate depression. Eur Neuropsychopharmacol **20:** 747-765

Katagiri M, Ganguli BN, Gunsalus IC (1968) A soluble cytochrome P-450 functional in methylene hydroxylation. J Biol Chem **243:** 3543-3546

Kerdpin O, Elliot DJ, Boye SL, Birkett DJ, Yoovathaworn K, Miners JO (2004) Differential contribution of active site residues in substrate recognition sites 1 and 5 to cytochrome P450 2C8 substrate selectivity and regioselectivity. Biochemistry **43:** 7834-7842

Kirmizibekmez H, Bassarello C, Piacente S, Celep E, Atay I, Mercanoglu G, Yesilada E (2009) Phenolic compounds from *Hypericum calycinum* and their antioxidant activity. Nat Prod Commun **4:** 531-534

Kitanov GM, Nedialkov PT (2001) Benzophenone *O*-glucoside, a biogenic precursor of 1,3,7-trioxygenated xanthones in *Hypericum annulatum*. Phytochemistry **57:** 1237-1243

Klein AP, Sattely ES (2015) Two cytochromes P450 catalyze S-heterocyclizations in cabbage phytoalexin biosynthesis. Nat Chem Biol **11:** 837-839

Klingauf P, Beuerle T, Mellenthin A, El-Moghazy SAM, Boubakir Z, Beerhues L (2005) Biosynthesis of the hyperforin skeleton in *Hypericum calycinum* cell cultures. Phytochemistry **66:** 139-145

Koh JJ, Qiu S, Zou H, Lakshminarayanan R, Li J, Zhou X, Tang C, Saraswathi P, Verma C, Tan DT, Tan AL, Liu S, Beuerman RW (2013) Rapid bactericidal action of alpha-mangostin against MRSA as an outcome of membrane targeting. Biochim Biophys Acta **1828:** 834-844

Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. Proteins **64:** 559-574

Kraus PF, Kutchan TM (1995) Molecular cloning and heterologous expression of a cDNA encoding berbamunine synthase, a C–O phenol-coupling cytochrome P450 from the higher plant *Berberis stolonifera*. Proc Natl Acad Sci U S A **92:** 2071-2075

Krieger E, Darden T, Nabuurs SB, Finkelstein A, Vriend G (2004) Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins **57:** 678-683

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins **77 Suppl 9:** 114-122

Krieger E, Nabuurs SB, Vriend G (2003) Homology modeling. Methods Biochem Anal **44:** 509-523

Krieger E, Vriend G (2014) YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics **30:** 2981-2982

Krohn K, Kouam SF, Kuigoua GM, Hussain H, Cludius-Brandt S, Florke U, Kurtan T, Pescitelli G, Di Bari L, Draeger S, Schulz B (2009) Xanthones and oxepino[2, 3-b]chromones from three endophytic fungi. Chemistry **15:** 12121-12132

Kubin A, Wierrani F, Burner U, Alth G, Grunberger W (2005) Hypericin–the facts about a controversial agent. Curr Pharm Des **11:** 233-253

Lamb DC, Lei L, Warrilow AG, Lepesheva GI, Mullins JG, Waterman MR, Kelly SL (2009) The first virally encoded cytochrome P450. J Virol **83:** 8266-8269

Lamb DC, Waterman MR (2013) Unusual properties of the cytochrome P450 superfamily. Philos Trans R Soc Lond B Biol Sci **368:** 20120434

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947-2948

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol **25:** 1307-1320

Lech K, Brent R (2001) Growing bacteria in liquid media. Curr Protoc Cytom **Appendix 3:** Appendix 3F

Lentz O, Feenstra A, Habicher T, Hauer B, Schmid RD, Urlacher VB (2006) Altering the regioselectivity of cytochrome P450 CYP102A3 of *Bacillus subtilis* by using a new versatile assay system. Chembiochem **7:** 345-350

Leuner K, Kazanski V, Muller M, Essin K, Henke B, Gollasch M, Harteneck C, Muller WE (2007) Hyperforin--a key constituent of St. John's wort specifically activates TRPC6 channels. FASEB J **21:** 4101-4111

Li Z, Rupasinghe SG, Schuler MA, Nair SK (2011) Crystal structure of a phenol-coupling P450 monooxygenase involved in teicoplanin biosynthesis. Proteins **79:** 1728-1738

Linde K, Berner MM, Kriston L (2008) St John's wort for major depression. Cochrane Database Syst Rev**:** CD000448

Lindstrom A, Ooyen C, Lynch ME, Blumenthal M, Kawa K (2014) Sales of herbal dietary supplements increase by 7.9% in 2013, marking a decade of rising sales: Turmeric supplements climb to top ranking in natural channel HerbalGram**:** 52-56

Linsmaier EM, Skoog F (1965) Organic growth factor requirements of *Tobacco* tissue cultures. Physiol Plant **18:** 100-127

Liu B, Falkenstein-Paul H, Schmidt W, Beerhues L (2003) Benzophenone synthase and chalcone synthase from *Hypericum androsaemum* cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases. Plant J **34:** 847-855

Liu C-J, Huhman D, Sumner LW, Dixon RA (2003) Regiospecific hydroxylation of isoflavones by cytochrome P450 81E enzymes from *Medicago truncatula*. Plant J **36:** 471-484

Liu H, Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol **8:** 91

Lu AY, Junk KW, Coon MJ (1969) Resolution of the cytochrome P-450-containing omega-hydroxylation system of liver microsomes into three components. J Biol Chem **244:** 3714-3721

Mahabusarakam W, Nuangnaowarat W, Taylor WC (2006) Xanthone derivatives from *Cratoxylum cochinchinense* roots. Phytochemistry **67:** 470-474

Mandel M, Higa A (1970) Calcium-dependent bacteriophage DNA infection. J Mol Biol **53:** 159-162

Martinez A, Hernandez-Marin E, Galano A (2012) Xanthones as antioxidants: A theoretical study on the thermodynamics and kinetics of the single electron transfer mechanism. Food Funct **3:** 442-450

Masters KS, Bräse S (2012) Xanthones from fungi, lichens, and bacteria: The natural products and their synthesis. Chem Rev **112:** 3717-3776

Matsumoto K, Akao Y, Kobayashi E, Ohguchi K, Ito T, Tanaka T, Iinuma M, Nozawa Y (2003) Induction of apoptosis by xanthones from mangosteen in human leukemia cell lines. J Nat Prod **66:** 1124-1127

Mattson MP (2004) Pathways towards and away from Alzheimer's disease. Nature **430:** 631-639

Mbwambo ZH, Kapingu MC, Moshi MJ, Machumi F, Apers S, Cos P, Ferreira D, Marais JP, Vanden Berghe D, Maes L, Vlietinck A, Pieters L (2006) Antiparasitic activity of some xanthones and biflavonoids from the root bark of *Garcinia livingstonei*. J Nat Prod **69:** 369-372

McDonald E (2008) Biosynthesis of isoquinolines. Chem Heterocycl Compd**:** 275-379

McLean KJ, Carroll P, Lewis DG, Dunford AJ, Seward HE, Neeli R, Cheesman MR, Marsollier L, Douglas P, Smith WE, Rosenkrands I, Cole ST, Leys D, Parish T, Munro AW (2008) Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J Biol Chem **283:** 33406-33416

Medina MA, Martinez-Poveda B, Amores-Sanchez MI, Quesada AR (2006) Hyperforin: More than an antidepressant bioactive compound? Life Sci **79:** 105-111

Miller AL (1998) St. John's wort (*Hypericum perforatum*): Clinical effects on depression and other conditions. Altern Med Rev **3:** 18-26

Mizutani M, Ohta D (1998) Two isoforms of NADPH:cytochrome P450 reductase in *Arabidopsis thaliana*. Gene structure, heterologous expression in insect cells, and differential regulation. Plant physiol **116:** 357-367

Mizutani M, Ohta D (2010) Diversification of P450 genes during land plant evolution. Annu Rev Plant Biol **61:** 291-315

Mizutani M, Ohta D, Sato R (1997) Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from *Arabidopsis* and its expression manner in planta. Plant Physiol **113:** 755-763

Mizutani M, Sato F (2011) Unusual P450 reactions in plant secondary metabolism. Arch Biochem Biophys **507:** 194-203

Molinar-Toribio E, González J, Ortega-Barría E, Capson TL, Coley PD, Kursar TA, McPhail K, Cubilla-Rios L (2006) Antiprotozoal activity against *Plasmodium falciparum* and *Trypanosoma cruzi* of xanthones isolated from *Chrysochlamys tenuis*. Pharm Biol **44:** 550-553

Morikawa T, Mizutani M, Aoki N, Watanabe B, Saga H, Saito S, Oikawa A, Suzuki H, Sakurai N, Shibata D, Wadano A, Sakata K, Ohta D (2006) Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in *Arabidopsis* and tomato. Plant Cell **18:** 1008-1022

Müller WE (2003) Current St John's wort research from mode of action to clinical efficacy. Pharmacol Res **47:** 101-109

Müller WE, Rolli M, Schafer C, Hafner U (1997) Effects of *Hypericum* extract (LI 160) in biochemical models of antidepressant activity. Pharmacopsychiatry **30 Suppl 2:** 102-107

Mullis KB (1994) The Polymerase Chain Reaction (Nobel Lecture). Angew Chem Int Ed Engl **33:** 1209-1213

Munro AW, Girvan HM, Mason AE, Dunford AJ, McLean KJ (2013) What makes a P450 tick? Trends Biochem Sci **38:** 140-150

Mustafa G, Yu X, Wade RC (2014) Structure and dynamics of human drug-metabolizing cytochrome P450 enzymes. *In* Drug Metabolism Prediction. Wiley-VCH Verlag GmbH & Co. KGaA, pp 75-102

Nahrstedt A, Butterweck V (2010) Lessons learned from herbal medicinal products: The example of St. John's wort. J Nat Prod **73:** 1015-1021

Nakatani K, Nakahata N, Arakawa T, Yasuda H, Ohizumi Y (2002) Inhibition of cyclooxygenase and prostaglandin E2 synthesis by gamma-mangostin, a xanthone derivative in mangosteen, in C6 rat glioma cells. Biochem Pharmacol **63:** 73-79

Narhi LO, Fulco AJ (1987) Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient monooxygenase induced by barbiturates in *Bacillus megaterium*. J Biol Chem **262:** 6683-6690

Neary JT, Bu Y (1999) *Hypericum* LI 160 inhibits uptake of serotonin and norepinephrine in astrocytes. Brain Res **816:** 358-363

Nelson D, Werck-Reichhart D (2011) A P450-centric view of plant evolution. Plant J **66:** 194-211

Nelson DR (2006) Cytochrome P450 nomenclature, 2004. Methods Mol Biol **320:** 1-10

Nelson DR (2009) The cytochrome P450 homepage. Hum genomics **4:** 59-65

Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC, Nebert DW (1996) P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenetics **6:** 1-42

Nelson DR, Ming R, Alam M, Schuler MA (2008) Comparison of cytochrome P450 genes from six plant genomes. Trop Plant Biol **1:** 216-235

Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative genomics of rice and *Arabidopsis*. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiol **135:** 756-772

Nurk NM (2011) Phylogenetic analyses in St. John's wort (*Hypericum*) inferring character evolution and historical biogeography. Free University of Berlin, Berlin, Germany

Nurk NM, Madrinan S, Carine MA, Chase MW, Blattner FR (2013) Molecular phylogenetics and morphological evolution of St. John's wort (Hypericum; Hypericaceae). Mol Phylogenet Evol **66:** 1-16

Ogliaro F, de Visser SP, Cohen S, Sharma PK, Shaik S (2002) Searching for the second oxidant in the catalytic cycle of cytochrome P450: a theoretical investigation of the iron(III)-hydroperoxo species and its epoxidation pathways. J Am Chem Soc **124:** 2806-2817

Omura T, Sanders E, Estabrook RW, Cooper DY, Rosenthal O (1966) Isolation from adrenal cortex of a nonheme iron protein and a flavoprotein functional as a reduced triphosphopyridine nucleotide-cytochrome P-450 reductase. Arch Biochem Biophys **117:** 660-673

Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J Biol Chem **239:** 2370-2378

Ondeyka JG, Dombrowski AW, Polishook JP, Felcetto T, Shoop WL, Guan Z, Singh SB (2006) Isolation and insecticidal/anthelmintic activity of xanthonol, a novel bis-xanthone, from a non-sporulating fungal species. J Antibiot **59:** 288-292

Ono E, Nakai M, Fukui Y, Tomimori N, Fukuchi-Mizutani M, Saito M, Satake H, Tanaka T, Katsuta M, Umezawa T, Tanaka Y (2006) Formation of two methylenedioxy bridges by a *Sesamum* CYP81Q protein yielding a furofuran lignan, (+)-sesamin. Proc Natl Acad Sci U S A **103:** 10116-10121

Onoue S, Seto Y, Ochi M, Inoue R, Ito H, Hatano T, Yamada S (2011) *In vitro* photochemical and phototoxicological characterization of major constituents in St. John's wort (*Hypericum perforatum*) extracts. Phytochemistry **72:** 1814-1820

Öztürk Y, Aydin S, Beis R, Başer KHC, Berberoğlu H (1996) Effects of *Hypericum perforatum* L. and *Hypericum calycinum* L. extracts on the central nervous system in mice. Phytomedicine **3:** 139-146

Park KH, Park YD, Han JM, Im KR, Lee BW, Jeong IY, Jeong TS, Lee WS (2006) Anti-atherosclerotic and anti-inflammatory activities of catecholic xanthones and flavonoids isolated from *Cudrania tricuspidata*. Bioorg Med Chem Lett **16:** 5580-5583

Peres V, Nagem TJ (1997) Trioxygenated naturally occurring xanthones. Phytochemistry **44:** 191-214

Peters S, Schmidt W, Beerhues L (1997) Regioselective oxidative phenol couplings of 2,3′,4,6-tetrahydroxybenzophenone in cell cultures of *Centaurium erythraea* RAFN and *Hypericum androsaemum* L. Planta **204:** 64-69

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res **29:** e45

Pfalz M, Mikkelsen MD, Bednarek P, Olsen CE, Halkier BA, Kroymann J (2011) Metabolic engineering in *Nicotiana benthamiana* reveals key enzyme functions in *Arabidopsis* indole glucosinolate modification. Plant Cell **23:** 716-729

Pfalz M, Vogel H, Kroymann J (2009) The gene controlling the *indole glucosinolate modifier1* quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. Plant Cell **21:** 985-999

Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol **272:** 51-64

Porter TD, Beck TW, Kasper CB (1990) NADPH-cytochrome P-450 oxidoreductase gene organization correlates with structural domains of the protein. Biochemistry **29:** 9814-9818

Poulos TL, Finzel BC, Howard AJ (1987) High-resolution crystal structure of cytochrome P450cam. J Mol Biol **195:** 687-700

Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S (2006) An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun **347:** 747-751

Ralston L, Yu O (2006) Metabolons involving plant cytochrome P450s. Phytochem Rev **5:** 459-472

Renault H, Bassard JE, Hamberger B, Werck-Reichhart D (2014) Cytochrome P450-mediated metabolic engineering: Current progress and future challenges. Curr Opin Plant Biol **19:** 27-34

Reutrakul V, Anantachoke N, Pohmakotr M, Jaipetch T, Sophasan S, Yoosook C, Kasisit J, Napaswat C, Santisuk T, Tuchinda P (2007) Cytotoxic and anti-HIV-1 caged xanthones from the resin and fruits of *Garcinia hanburyi*. Planta Med **73:** 33-40

Rittle J, Green MT (2010) Cytochrome P450 Compound I: Capture, characterization, and C–H bond activation kinetics. Science **330:** 933-937

Ro D-K, Ehlting J, Douglas CJ (2002) Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. Plant Physiol **130:** 1837-1851

Robson NKB (2003) *Hypericum* botany. *In* E Ernst, ed, *Hypericum* - The Genus *Hypericum*. Taylor and Francis, London, New York, pp 1-22

Rukachaisirikul V, Phainuphong P, Sukpondma Y, Phongpaichit S, Taylor WC (2005) Antibacterial caged-tetraprenylated xanthones from the stem bark of *Garcinia scortechinii*. Planta Med **71:** 165-170

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol **234:** 779-815

Schalk M, Croteau R (2000) A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (-)-limonene hydroxylase from a C6- to a C3-hydroxylase. Proc Natl Acad Sci U S A **97:** 11948-11953

Schmidt W, Beerhues L (1997) Alternative pathways of xanthone biosynthesis in cell cultures of *Hypericum androsaemum* L. FEBS Lett **420:** 143-146

Schmidt W, Peters S, Beerhues L (2000) Xanthone 6-hydroxylase from cell cultures of *Centaurium erythraea* RAFN and *Hypericum androsaemium* L. Phytochemistry **53:** 427-431

Schuler MA, Rupasinghe SG (2011) Chapter 5 - Molecular and Structural Perspectives on Cytochrome P450s in Plants. *In* K Jean-Claude, D Michel, eds, Advances in Botanical Research, Vol Volume 60. Academic Press, pp 263-307

Schulz V (2006) Safety of St. John's wort extract compared to synthetic antidepressants. Phytomedicine **13:** 199-204

Schwarz H, Liu B, Peters S, Barillas W, Beerhues L (2009) Purification, cDNA cloning and functional expression of NADPH-cytochrome P450 reductase from *Centaurium erythraea* cell cultures. Plant Biol **11:** 300-306

Seifert A, Pleiss J (2009) Identification of selectivity-determining residues in cytochrome P450 monooxygenases: A systematic analysis of the substrate recognition site 5. Proteins **74:** 1028-1035

Sethiya NK, Mishra SH (2014) Investigation of Mangiferin, as a Promising Natural Polyphenol Xanthone on Multiple Targets of Alzheimer's Disease. J Biol Act Prod Nat **4:** 111-119

Shadid KA, Shaari K, Abas F, Israf DA, Hamzah AS, Syakroni N, Saha K, Lajis NH (2007) Cytotoxic caged-polyprenylated xanthonoids and a xanthone from *Garcinia cantleyana*. Phytochemistry **68:** 2537-2544

Sirim D, Widmann M, Wagner F, Pleiss J (2010) Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. BMC Struct Biol **10:** 34

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res **33:** W244-248

Stadler R, Zenk MH (1993) The purification and characterization of a unique cytochrome P-450 enzyme from *Berberis stolonifera* plant cell cultures. J Biol Chem **268:** 823-831

Stegmann E, Pelzer S, Bischoff D, Puk O, Stockert S, Butz D, Zerbe K, Robinson J, Sussmuth RD, Wohlleben W (2006) Genetic analysis of the balhimycin (vancomycin-type) oxygenase genes. J Biotechnol **124:** 640-653

Stevens PF (2007) Clusiaceae-Guttiferae. *In* K Kubitzki, ed, The Families and Genera of Vascular Plants, Flowering Plants · Eudicots, Vol 9. Springer Berlin Heidelberg, pp 48-66

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol **30:** 2725-2729

Thiede HM, Walper A (1994) Inhibition of MAO and COMT by *Hypericum* extracts and hypericin. J Geriatr Psychiatry Neurol **7 Suppl 1:** S54-56

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:** 4673-4680

Tocci N, Simonetti G, D'Auria FD, Panella S, Palamara AT, Valletta A, Pasqua G (2011) Root cultures of *Hypericum perforatum* subsp. *angustifolium* elicited with chitosan and production of xanthone-rich extracts with antifungal activity. Appl Microbiol Biotechnol **91:** 977-987

Treco DA, Winston F (2001) Growth and manipulation of yeast. Curr Protoc Mol Biol **Chapter 13:** Unit13.12

Urlacher VB, Eiben S (2006) Cytochrome P450 monooxygenases: Perspectives for synthetic application. Trends Biotechnol **24:** 324-330

Vieira LM, Kijjoa A (2005) Naturally-occurring xanthones: Recent developments. Curr Med Chem **12:** 2413-2446

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem **25:** 1157-1174

Wang JN, Hou CY, Liu YL, Lin LZ, Gil RR, Cordell GA (1994) Swertifrancheside, an HIV-reverse transcriptase inhibitor and the first flavone-xanthone dimer, from *Swertia franchetiana*. J Nat Prod **57:** 211-217

Wang Y, Shi J-G, Wang M-Z, Che C-T, Yeung JHK (2007) Mechanisms of the vasorelaxant effect of 1-hydroxy-2, 3, 5-trimethoxy-xanthone, isolated from a Tibetan herb, *Halenia elliptica*, on rat coronary artery. Life Sci **81:** 1016-1023

Wang Y, Xia Z, Xu JR, Wang YX, Hou LN, Qiu Y, Chen HZ (2012) α-mangostin, a polyphenolic xanthone derivative from mangosteen, attenuates β-amyloid oligomers-induced neurotoxicity by inhibiting amyloid aggregation. Neuropharmacology **62:** 871-881

Wang Z, Gorski JC, Hamman MA, Huang SM, Lesko LJ, Hall SD (2001) The effects of St John's wort (*Hypericum perforatum*) on human cytochrome P450 activity. Clin Pharmacol Ther **70:** 317-326

Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: A success story. Genome Biol **1:** REVIEWS3003

Wezeman T, Brase S, Masters KS (2015) Xanthone dimers: A compound family which is both common and privileged. Nat Prod Rep **32:** 6-28

Win T, Htwe TT, Shwe HH, Heilmann J (2012) Lavandulyl flavanones from the stems of *Hypericum calycinum* L. Chem Biodivers **9:** 1198-1204

Winter DK, Sloman DL, Porco JA, Jr. (2013) Polycyclic xanthone natural products: Structure, biological activity and chemical synthesis. Nat Prod Rep **30:** 382-391

Woithe K, Geib N, Zerbe K, Li DB, Heck M, Fournier-Rousset S, Meyer O, Vitali F, Matoba N, Abou-Hadeed K, Robinson JA (2007) Oxidative phenol coupling reactions catalyzed by OxyB: A cytochrome P450 from the vancomycin producing organism. Implications for vancomycin biosynthesis. J Am Chem Soc **129:** 6887-6895

Wurdack KJ, Davis CC (2009) Malpighiales phylogenetics: Gaining ground on one of the most recalcitrant clades in the angiosperm tree of life. Am J Bot **96:** 1551-1570

Xu J, Wang X-y, Guo W-z (2015) The cytochrome P450 superfamily: Key players in plant development and defense. J Integr Agric **14:** 1673-1686

Yang CQ, Lu S, Mao YB, Wang LJ, Chen XY (2010) Characterization of two NADPH: Cytochrome P450 reductases from cotton (*Gossypium hirsutum*). Phytochemistry **71:** 27-35

Yang LJ, Chen Y (2013) New targets for the antitumor activity of gambogic acid in hematologic malignancies. Acta Pharmacol Sin **34:** 191-198

Yim G, Thaker MN, Koteva K, Wright G (2014) Glycopeptide antibiotic biosynthesis. J Antibiot **67:** 31-41

Zawaira A, Ching LY, Coulson L, Blackburn J, Wei YC (2011) An expanded, unified substrate recognition site map for mammalian cytochrome P450s: Analysis of molecular interactions between 15 mammalian CYP450 isoforms and 868 substrates. Curr Drug Metab **12:** 684-700

Zhang H-Z, Kasibhatla S, Wang Y, Herich J, Guastella J, Tseng B, Drewe J, Cai SX (2004) Discovery, characterization and SAR of gambogic acid as a potent apoptosis inducer by a HTS assay. Bioorg Med Chem **12:** 309-317

Zhang HY (2005) One-compound-multiple-targets strategy to combat Alzheimer's disease. FEBS Lett **579:** 5260-5264

Zhang Y, van Dijk AD, Scaffidi A, Flematti GR, Hofmann M, Charnikhova T, Verstappen F, Hepworth J, van der Krol S, Leyser O, Smith SM, Zwanenburg B, Al-Babili S, Ruyter-Spira C, Bouwmeester HJ (2014) Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis. Nat Chem Biol **10:** 1028-1033

Zhao J, Liu W, Wang JC (2015) Recent advances regarding constituents and bioactivities of plants from the genus *Hypericum*. Chem Biodivers **12:** 309-349

VIII. Appendix

1. Sequences

1.1. Assembled contigs from *H. calycinum* subtractive library encoding CYPs

>Contig21
CGATTAGCGTGGTCGCGGCCGAGGTACTTCGAGAGCGTCTCCAAGAAACTTGGCCCGGTC
TTGCTCTTCTGGATGGGCAAGGTCCCCGGCCGCATGGCGGCCGCGGGAATCGATTAGCGT
GGTCGCGGCCGAGGTACAAGGGGATCGAGAAGAGCATGATCCAGCTGGCCGAGAAGAGGG
AGGTCTTCCTTAACGCCCTGGTCGCCGAGGTCAGAGCCAGGAGGGCCGCCGGCGGGAAGA
CGACAGTTGGCGAGAACGGCGAAGGGGTCTCTGTGATCGATATCATCCTCAACTTGCAGG
AGACAGACCCCGAGTTTTACACCGATGCCATCGTCAAGGGTATTGTTATGATGATGTTTA
TCGCCGGGACCGAGACGTCGACCGCCACGTTGGAGTGGGGGTTCACTCTCCTAATGCAAC
ACCCAGAGAAGATGCGCAAACTACAAGCCGAGATCGACAGCGTCGTAGGAGACAGCCGCT
TCGTGAGCGAAGCCGACTTCGCCAATCTTCCCTACCTAAGGGCCGTCGTGAAGGAGACAC
TAAGGCTGTACCTCGCCCGGGCCCACGCTAGAACTCCCCCACTTCTCTAACCAAGCCTGC
ACCGTCGGAGGCTACGACATTCCCAATGGAACCATGCTCTTCGCCAACTTGTGGGCAATG
CAGAGAGACCCTAAAGTATGGGACGAGCCAGACGCATTCAAGCCCGAGAGGTTCGAGGAA
TACGACCGCGAAGGCGGGTACCGGTCTTTTTCGCCAGACAGGCAAGAGCGCTTCGTCAAC
AGATGGATTAGCTCGCTATCCGACCCCCGTGTCACCCATGAAATCAGAGCCATCTGGGTC
ACCTACTGGTATCAGGCTGACAAATCTCTTGGTCAGAAGCTCGCATCTCGTCTCAATGTG
AGGCCAAACATGTGAATTTGATGCCGGCCGGGCACTTTCTGCTGCCGAGACATGAGAGAA
AACGTTGGAAGGGGCTGTTATATGCAATGAATCTACATCAGAGAATAACATCTAAATGTG
TGTGTGTTCTGTGTTGTAATACTGAAGCATAGTTTCGTGTGTTGCTCCCCCTAATTAAAC
ATTCTTATCTAATAAACGGGTTTGTACCTGCCCGGGCGGCCGCTCGACCTTTGAGGAGTG
CTTCA

>Contig41
GCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACACCACCGATCCTGC
GAAGGCGGCAGCAGCAGCTGCCAGCAGTATGAAAGGTGGTGCCTCGCCTTATGTCCAGTT
TGCTGCTGCTCAGTCATCTGGAAAGCCGCATCAGCTTGTAACCGCAGGGTTTCCCTATGT
TCCGGCGGCTGTTCAAGTGAAACAAGCTGATCAGAAGCAACCCGCTGGTGAGTAGAAGTT
CTCCGTAGCGCTTACATGTTGAACCCCATGGCGGCGCGGGAATCGAGGTCGAGCGGCCGC
CCGGGCAGGTACTTCCGCAGGCGGCCGGCGAGGTGCCGGATCTCCTCCTCCCGGATGAAG
GCCGACCTCTGGAGGGCCTTGATGGAGAAGATCTCCGAGAAGGTGAGCCGACGGAGGCTC
TTCCAGATGGGGCCGTGGGGGGCCCACACGAGGAAGGTGTAGTCATAGGTGAGATGCTCG
CCGGACATGGACTTGGGGCGGTCGGCAAAGACGACGTCGTTCTTGGTGAAGCACTCCTCA
AAGGTCGAGCGGCCGCCCGGGCAGGTACGACCGGGCGTGAGAGGCACTCGGTCCTCCGGA
TTTTCAAGGGCCGCCGGGGGCGCACCGGACACCACGCGACGTGCGGTGCTCTTCCAGCCG
CTGGACCCTACCTCCGACTGAGTCGTTTCCAGGGTGGGCAGGCTGTTAAACAGAAAAGAT
AACTCTTCCCGAGGCCCCCGCCGACGTCTCCGGACTCCCTAACGTTGCCGTCAGCCGCCA
CGTCCCGGTTCAGGAATTTAACCCGATTCCCTTTCGAAGTACCTCGGCCGCGACCACGCT
AATCGATTCCCGCGCCG

>Contig50
GGCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACTTCGAGAGCGTCTCCAAGAAAC
TTGGCCCGGTCTTGCTCTTCTGGATGGGCAAGGTCCCCGTCCTCGTTGTCTCCTCCCCGG
AGGTCGAGCGGCCGCCCGGGCAGGTACTTAAACAAATCACTGCCTCTCTATATCTTCAAC
TCGAGCTCCCTCACCAGAGCTTCGAGACCTGGTTACAAAATTCCTGAAGGAGCCCAGCCT
TGAGCCCGACCCATCAACCTTCTCCAATTGGGCCGTATTGGGCCCGTTATTATCCTCCGG
GGCTGAAGGTTGGACCCTAGGCCCATTATGGCTTCCTTGTTGGGCCTTCGGCTCTGCCTC
GGTTCGACATATCGGGCACGTCGAGTGCGAGCCCAACCACATGTCCACGCACTCGACGTG
AAACATATGGTTACAATTTGGCAAAACTCTTACCACTTCCTCGTCCTCTATTCCCATTAA
GCAAACCGAGCACTCCACTTGGTCGCCCCGGCCGTACCTCGGCCGCGGCGATTCGAGCGG
CCGCCCGGGCAGGTACTTCCGCAGGCGGCCGGCGAGGTGGCGGATCTCCTCCTCCCGGAT
GAAGGCCGACCTCTGGAGGGCCTTGATGGAGAAGATCTCAGAGAAGGTGAGACGACGGAG
GCTCTTCCAGATGGGGCCGTGGGGGGCCCACACGAGGAAGGTGTAGTCATAGGTGAGGTG
CTCGCCGGACATGGACTTGGGGCGGTCGGCGAAGACGACGTCGTTCTTGGTGAAGCACTC
CTCGAAGGCCTCCGGAGAGGAGACAACGAGGACGGGGACCTTGCCCATCCAGAAGAGCAA
GACCGGGCCAAGCTTCTTGGAGACGCTCTCGAAGTACCTCGGCCGCGACCACGCTAATCG
ATTCCCGCGGCCGCCATGCGGCCGGAGC

>Contig56
TCGATTGGTACCTCAGTATGGTCGTAAGAGAAACTTTCAGACTCCATCCACCAGTGCCAT
TGCTACTCCCCCACGCTGCAAGAGAAGAATGTGACATCAATGGCTTTTACATTCCGAAGA
ATTCGCATGTTATCTTCAACGCATGGGCGATCGGCCGCGACCCTGGCGCTTGGGTCGACC
CGGAGAAGTTTTGGCCGGAGAGATTCAGAGGAAGTGAGGTGGACGTCCGTGGACGTGATT
TCCAGCCGATCCCATTTGGCTCTGGACGTAGAGGGTGTCCTGGGATGCAATTAGCCCTCG
TATTTGTGCACTTGGTTTTGGCCAATTTGGTGCATTCGTTTGAGTGGGAGCTGCCGGGTG
GGATGTTGCCTCATGAATTGGATATGACAGAGGAGTACCTCGGCCGCGACCACGCTA

>Contig59
TCCGGCCGCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACTTCGAGA
GCGTCTCCAAGAAACTTGGCCCGGTCTTGCTCTTCTGGATGGGCAAGGTCCCCGTCCTCG
TTGTCTCCTCCCCGGAGGCCTTTGAGGAGTGCTTCACCAAGAACGACGTCGTCTTTGCCG
ACCGCCCCAAGTCCATGTCCGGCGAGCATCTCACCTATGACTACACCTTCCTCGTGTGGG
CCCCCCACGGCCCCATCTGGAAGAGCCTCCGTCGGCTCACCTTCTCGGAGATCTTCTCCA
TCAAGGCCCTCCAGAGGTCGGCCTTCATCCGGGAGGAGGAGATCCGGCACCTCGCCGGCC
GCCTGCGGAAGTACCTGCCCGGGCGGCCGCTCGAATCGATTCCCGCGGCCGCCATGCGGC
CGTTTGAGGGCCAAGTTTGGGTTCTTTTGCTACTTCACGGCGGTGGCCATCCCGTTCTTG
AGCAGCCTGGCCGGTCTCATCGGAGGAATTGCGTTGCCGGTGACGTTCGCTTACCCGTGC
TTCATGTGGCTCAGGATCAAGAAGCCCAAGGTTTACGGCCCAATGTGGTTCCTCAATTGG
ACTCTTGGGCTCTCCGGTTTGGGCCTGAGTGGGGCCGCCATTGCAGCTAGCATTTATGTC
ATAATTGACACCGGAATTAACTTTAGCTTTT

>Contig62
GGCCGCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACTCGACTGTCT
GATTTGAACGAGAAGTTTCGCTAGTTCTTTGATAAGATCAAATATTCTTTTGCCTTCCTC
GTAACTGCTTCCAAATGCTGTTCTAGAGATTACATCGCAGGTTAGTTCTTAAAGATGCGG
CCAAACATCTACCTCGCATGATCCTTCCTCGGAGGCCGATCGCTCCCATTTACCGATCAT
GTCACTACAACATCGACTAAATACGGGCATCATTACCTTCAATTTTTCAACATGAAATGC
TGGATTGATGATCCTTCGGTGCTTGGCCCATTTTTCCCCTTCATAGCTCGCGAGGTAATC
GATTAGCGGCCGCCCGGCCGAGGTACGGACAAGGGGAATCCGACTGTTTAATTAAAACAA
AGCATTGCGATGGTCCCTGCGGATGTTAACGCAATGTGATTTCTGCCCAGTGCTCTGAAT
GTCAAAGTGAAGAAATTCAACCAAGCGCGGGTAAACGGCGGGAGTAACTATGACTCTCTT
AAGGTAGCCAAATGCCTCGTCATCTAATTAGTGACGCGCATGAATGGATTAACGAGATTC
CCACTGTCCCTGTCTACTATCCAGCGAAACCACAGCCAAGGGAACGGGCTTGGCAGAATC
AGCGGGGAAAGAAGACCCTGTTGAGCTTGACTCTAGTCCGACTTTGTGAAATGACTTGAG
AGGTGTAGGATAAGTGGGAGCCGGAAACGGCGAAGGTGAAATACCACTACTTTTAACGTT
ATTTTACTTATTCCGTGAATCGGAGGCGGGGCATCGCCCCTTCTTTTGGACCTAAGGTCG
CTTCGGCGGCTGATCCGGGCGGAAGACATTGTCAGGTGGGGAGTTTGGCTGGGGCGGCAC
ATCTGTTAAAAGATAACGCAGGTGTCCTAAGATGAGCTCAACGAGAACAGAAATCTCGTG
TGGAACAAAAGGGTAAAAGCTCGTTTGATTCTGATTTCCAGTACCTGCCCGGGCGGCCGC
TCGAACTCCGGGGAGGAGACAACGAGGACGGGGACCTTGCCCATCCAGAAGAGCAAGACC
GGGCCAAGTTTCTTGGAGACGCTCTCGAAGTACCTCGGCCGCGACCACGCTAATCGATTC
CCGCGGCCGCCATGCG

>Contig86
CATGCTCCGGCGCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACTTC
GAGAGCGTCTCCAAGAAGCTTGGCCCGGTCTTGCTCTTCTGGATGGGCAAGGTCCCCGTC
CTCGTTGTCTCCTCTCCGGAGGTCGAGCGGCCGCCCGGGCAGGTACGGCCCGGTGATCTC
CTTCAAGTTCGGGGCCCACCCGGCCGTGATCGTCTCGTCCCTCCCTGCGGCAGAGGAATG
CTTCACGAAGAACGACGTCGTCCCGGCCAACCGGCCCGTGTTCATGACCCTCTTCAGGCA
CCTAAACTACAACTACAGCACCATCTCCGCCGCACCCTATGGCGACCACTGGCGGAACCT
TCGTCGCATCAGCGCTACAGACATATTCTCCCCCCACCGCCTCAGCGTATTCCTCAGCAC
CCGACAGGAAGAGGTCCTGCGCATGCTGAAGAAATTGGAATGGCAATGGACAGAATCCGG
TTCAGACTCCGCCAAGGTCAACATCAAACCGTTGCTCTCGGACCTGACGTTCAACGTGAT
CATGAGAATGGTCGCGGGGAAGAGGTACCTCGGCCGCGACCACGCTA

>Contig103
GCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACCACCGGCCCCACTC
TCCCTCCCCCACTTCTCTAACCAAGCCTGCACCGTCGGAGGCTACGACATTCCCAATGGA
ACCATGCTCTTCGCCAACTTGTGGGCAATGCAGAGAGACCCTAAAGTATGGGACGAGCCA
GACGCATTCAAGCCCGAGAGGTTCGAGGAATACGACCGCGAAGGCGCGTACCTGCCCGGG
CGGCCGCTCGACCTTTGAGGAGTGCTTCACCAAGAACGACGTCGTCTTTGCCGACCGCCC
CAAGTCCATGTCCGGCGAGCATCTCACCTATGACTACACCTTCCTCGTGTGGGCCCCCCA
CGGCCCCATCTGGAAGAGCCTCCGTCGGCTCACCTTCTCGGAGATCTTCTCCATCAAGGC
CCTCCAGAGGTCGGCCTTCATCCGGGAGGAGGAGATCCGGCACCTCGCCGGCCGCCTGCG
GAAGTACCTGCCCGGGCGGCCGCTCCCTGGCCGCTGCTGGTGTTGCCGGATATGCCATTT
ACAAATACAGAATCAGGACATACATGGACTCGGAGATACGCGCTATAATGGCACGGTACC
TCGGCCGCGACCACGCTAATCGATTCCCGCGCCGGCCATGCGGCCGGAGCATG

>Contig110
GCCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACCCGTTCTGAGTTG
ACTGTTCGACGCCCGGGGGAAGGACCCCGAAGGGACCATTCCCAGTCCGTCCCCCGGCCG
GCACGCGACGACCCGCTCTCGCCGCGGAAGCAGCTCGAGCAGTCCACCGACAGCCGACGG
GTTCGGGACTGGGACCCCCGGACCCAGCCCTCAGAGCCAATCCTTTTCCCGAGGTTACGG
ATCCATTTTGCCGACTTCCCTTGCCTACATTGTTCCATTGGCCAGAGGCTGTTCACCTTG
GAGACCTGATGCGGTTATGAGTACGACCGGGCGTGAGAGGCACTCGGTACCTGCCCGGGC
GGCCGCTCGACCTTCGAGGAGTGCTTCACCAAGAACGACGTCGTCTTCGCCGACCGCCCC
AAGTCCATGTCCGGCGAGCACCTCACCTATGACTACACCTTCCTCGTGTGGGCCCCCCAC
GGCCCCATCTGGAAGAGCCTCCGTCGTCTCACCTTCTCTGAGATCTTCTCCATCAAGGCC
CTCCAGAGGTCGGCCTTCATCCGGGAGGAGGATATCCGCCACCTCG

>Contig158
CAGCTCCTCTTGGACTTTGGTTAGGACTCTTGGGGTGTTGATGACCTCTGCCATGGCTAA
CTCGATTATGTTCGTGGACGTGTCTGTTCCACCCACGATCATGTCCATAAGCAATGCCTT
GACCTGGATCAAGGAGAGCTTGGCCTTGCCGACAACCTCTTCATCCTCCAAGTTCAACAA
AAACTGCAAAAAGTCTGTATTCTCCTCCCCCTTCCCGATCGGCGTCCATCCTCCGGCGCC
GATCGATCATCCTCTCAAACATTGCGTCCATCTTCCCGGTGATCCTTAACATCCTCCGGT
GGAGTCCCTGGGGGTCAAACCGCACCATCCACGGGAAAAAGTCCGCCAAATTCGGCTTCA
CCATCAGCGCCGTCATTTCCACGACCGCCGCCCGGAACTCCCTCCCGACGCTCTCTCTCT
CCTCGTCGCCATGCACCGTCCCGGCCCAAAGCATGCTCGTGATCATGTTTATTATCGTCA
CCTG

>Contig181
ATGCTCCGGCCGCATGGCGGCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACGTGG
GAGGCTATACCATCCCGAAGGGCACGAGGGTCCTCGTCAATGCGTGGGCCATCCACAGGG
ACCCGTCAATATGGGATGACCCTCTGGAGTTTAGGCCCGAGAGGATGATGGATGGAAAGT
GGGACTACAACGGGAAGGACTTGAGCTATATCCCATTCGGGTCCGGGAGGAGGAGCTGTG
TCGGTGTCGCCATGGCCGAGAGGATGTTTATGTACCTCGGCCGCGACCACGCTA

>Contig199
CGATTAGCGTGGTCGCGGCCGAGGTGTCCCCGTCCTCGTTGTCTCCTCCCCGGAGGTCGA
GCGGCCGCCCGGGCAGGTACGCCCCCTGAGGTGGGCGAGCAGTGGTAAGTTCGAGAAGAG
AGTGAAGTCGCTTGGCGAGAGGACCGACAAGTTCTTGCAAGGGTTAGTTGACGAGCACCG
GAACAAGGAGAAGAGTCAAGAGGATGCGACCAACGTTACGATGATTGACCACTTGCTTTC
TTTGCAAGAGACTCAACCACAGGATTACTCTGATGTCACCATCAAAGGGCTTGTCCTGAT
AATGTTGCTGGCGGGAACTGATACATCAGCAGTGACGTTAGAGTGGGCTATGTCCAATCT
GGTTAACCACCCGAATGTCTTGGACAAGGCAAGGAAAGAACTGGACGCTCGAATCGGACA
AGATCGGCTCATCGAAGAAGACGATCTTCCTCATCTTTCTTACCTCCATAACATTATATC
AGAGACACTACGACTGTACCTCGCCCGCGACCACGCTAAACCATGACCGAGGCCGCCATG
CTATCGTTCCCACTTCATTCGAAGCTCTCGAGGGAGCAATCAAGGAAACATTTGATAAGC
TGGTTGAAGCGGGAAATATTACCCCCATAAAGGAGATCACACCTCCCCCTATTCCTGAGG
ACCTCAACAGTGCAATTAAGAGTGGCAAAGTCCGGGCTCCTACTCATATCATCTCT

>Contig219
GGCGCCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACCTCGAGAGCG
TCTCCAAGAAACTTGGCCCGGTCTTGCTCTTCTGGATGGGCAAGGTCCCCGTCCTCGTTG
TCTCCTCCCCGGAGGCCTTTGAGGAGTGCTTCACCAAGAACGACGTCGTCTTTGCCGACC
GCCCCAAGTCCATGTCCGGCGAGCATCTCACCTATGACTACACCTTCCTCGTGTGGGCCC
CCCACGGCCCCATCTGGAAGAGCCTCCGTCGGCTCACCTTCTCGGAGATCTTCTCCATCA
AGGCCCTCCAGAGGTCGGCCTTCATCCGGGAGGAGGAGATCCGGCACCTCGCCGGCCGCC
TGCGGAAGTACCTGCCCGGGCGGCCGCTCGACCTTCTCCCCTCTTGGTGTCTCCGACTGG
AACTCCCTCTTCTACTCCATCCACCCCGGCGGCCGTGGCATCATCGACGGCGTTGCCGGC
AACCTAGGGATCAAGGACGAGAACCTTGTGGCCACCAGGCACGTCCTCGGCGAGTACCTC
GGCCGCGACCACGCTA

1.2. Core fragments (ESTs) encoding CPR

>zwsca0_0001_G06.ab1 CHROMAT_FILE: zwsca0_0001_G06.ab1 PHD_FILE: zwsca0_0001_G06.ab1.phd.1 CHEM: term DYE: big TIME: Tue Jun 29 11:23:24 2010
GAGAAGGCGATAACCAGCTCCGAAAGTGCGCCACATTCAAGGAAATTGTT
TAGTTCATCTTCGTAGATGAAATCCATCCTGCGATTTCTGCAGCCGAAAA
ACAAAACGGCTGGTCCCAGTTCTGCTCCGGATTCTTTCAGGGCTAACCTT
TCCTGCAGGAAACCCCTGAATGGAGCCAAGCCAGTACCTGCCCGGGCGGC
CGCTCGAA

>zwsca0_0015_G08.ab1 CHROMAT_FILE: zwsca0_0015_G08.ab1 PHD_FILE: zwsca0_0015_G08.ab1.phd.1 CHEM: term DYE: big TIME: Tue Jun 29 11:24:24 2010
GCCATGGCGGCCGCGGGAATCGATTAGCGTGGTCGCGGCCGAGGTACTGG
CTTGGCTCCATTCAGGGGTTTCCTGCAGGAAAGGTTAGCCCTGAAAGAAT
CCGGAGCAGAGCTGGGACCAGCCGTTTTGTTTTTCGGCTGCAGAAATCGC
AGGATGGATCTCATCTACGAAGATGAACTAAACAATCGAGCGGCCGCCCG
GGCAGGTACCTGAGTCTCCATTTTGCCGGAACTAATTGCTATAAAGATGT
ATAACTTGGCTAGTTCGGGCATTCAACTTGGCAATTATAGTTGTCTTCAT
CAAGGAGGGAGCTAAGTCTCAACTCATAGTTGAGGTGGAGCGAAATCTCT
CGACCTTCTGCCAATGTAGGGAGATGCACACCCTGATAATATGGAGTGTG
GTCTCTTCGCGAATGGATGGTCTAGGAGCTGAGCAGCAGTGGGACGAGCA
CTGGGATCAACTTTCAGGCATTGCAGGATAAAATCCCTTGCATCAGTTGA
AAGAGATTCAGGAACAGATGGTGGTATGCCTTTGCCAATTCTGTATATAG
CCTGAAAAGAATCTAATTCCGAATAGGGAATCTGACTTGTTAACATCTCT
AACACAGTGCATCCAAGGCTCCATATATCTGCAGGAAGCCCATACCCTTG
GTTCTTTCGGTTCACAACCTCAGGAGCCATCCAGCATGCAGTACCTCGGC

1.3. Sequences from the *H. perforatum* transcriptome (MPGR)

1.3.1. Locus 416

>hpa_locus_416_iso_3_len_1885_ver_2
CACGACGCTCTTCCGATCTGTGAATTATGAGCACACTGAACGTCGGTACACAGCTCCATA
AATTTTTTTCGACAAGTACACATTATTATATATGTAAAAAAAGTATCCAGTTTCAAGGGA
TTGAAGTTTGGACCACATCTCAACTTACAGATCCCAAAATTAAGGAAGCATTATTGGATA
TTGCGTAGAAAGAAAAAATAAAATTAAACAAATATGCCAGCGCGAGAGAGGTTTAGAGCT
GGGAGAGGAGGCCAGCCATCTTGGTGCGAGGCGTGAAGTCAGCCTCCAGTGGCTTCTTCC
TTGACAGACTCGCGGTACCACGCGCGTGGCTCATGTCCTCGTCCATGCCGAGGTGGCCCC
AGTCGAAGCACTGCATAAGAATTCCAATCCCCAGGGACGAAATGTGGGTCCCCATCCCTC
CGCCGGGGCAAGACCTCCTCCCGACACCGAACGGCGCGTACTGGTACCCTCCTTCGCGGT
CGTACTCCTCGAACCTCTCGGGCTTGAACGCGTCGGGCTCGTCCCATACTTTGGGGTCTC
TCTGCATTGTCCACAGGTTGGCGAAGAGCATGGTTCCCTTGGGAATGTCGTAGCCTCCGA
CGGTGCAGGCTTGGTTGGAGAAGTGGGGGAGGGAGAGTGGGGCCGGCGGGTACATCCTGA
GCGTCTCCTTCACGACCGCCCTTAGGTAGGGAAGGTTGGCGAAGTCGGATTCGCACACGA
AGCGGCTGTCTCCGACGACGCTGTCGATCTCGGCTTGTAGTTTCCGCATCTTCTCAGGGT
GTTGCATCAGCAGTGTGAACGCCCACTCCAACGTGGCGGTCGACGTCTCGGTCCCGGCAA
TAAACATCATCATAACAATTCCCTTGACGATGTCATCCGTGTAAAACTCAGGGTCTGTCT
CTTGCAAGTTGAGGATGATATCGATAACGGAGACCCCTTCGCCGTTCTCGCCAACAGTGG
TCTTCCCGGCGGCCGCCCTCCTGGCTCTGACCTCGGCGACAAGGCCGTTGAGGAAGACCT
CCCTCTTCTCGGCCAGCTGGATCATGCTCTTCTCGATCCCCTTGTAACCGATCAGCCTCA
TCACCGGGATGAAGTCGCATATGGTCATGGAGACGATAGGAAAGAAGAGATCCTTGAACT
GCTGGAGCAGCCGCTTCTCCTCCTTGGTGTCCTCGTCGACGACCGGCACGTGCCTCTTCC
CGGCCGCCACCCGCATGATCACGCTTGACGTCAGCAGGGCGAACATGTGCTTCATGTCCA
CCTTGGAGTTGTTCCCTTTGACGGAGTACTTGAGAAGGCGGCCGGCCAGGTGGCGGATCT
CCTCCTCCCGGATGTAGGCGGACCTCTGGAGGGCCTTGATGGAGAAGATCTCGGAGAAGG
TGAGACGTCGGAGGCTCTTCCAGATGGGGCCGTGCGGGGCCCACACGAGGAAGGTGTAGT
CGTAGGTGAGGTGCTCACCTGACATGGACTTCGGGCGGTCGGCGAAGACGACGTCGTTCT
TTGTGAAGCACTCCTCGAAGGCCTCCGGGGAGGAGACGACGAGGACGGGGACCTTGCCCA
TCCAGAAGAGAAGGACTGGGCCGAGCTTCTTGGAGACGCTCTCGAAGTACTGGTAGAGCG
GGGGCTTGAGGACGGGGAGGTGTCCAACGATGGGATATGACCGTGGGCCCGGCGGGTAGT
TGAGCTTCTTACCGAAGAGAACCTTGTAGCCGATGATGACGGAGACAAGAACGGCGACGT
ACAGATACAAGTCCTCCATTTTATTTTTTTTTTCTTAAAGGAGTTGCTCTCTCTTCTCTA
TTTCACAAGACGACGACGACGACTTCTTTTTTTTCTGATATGTATTTTGAGGGTTGTGGA
ATAGAGAAGAGAGAGCAACTCCTTT

>hpa_locus_416_iso_3_len_1885_ver_2
MEDLYLYVAVLVSVIIGYKVLFGKKLNYPPGPRSYPIVGHLPVLKPPLYQYFESVSKKLG
PVLLFWMGKVPVLVVSSPEAFEECFTKNDVVFADRPKSMSGEHLTYDYTFLVWAPHGPIW
KSLRRLTFSEIFSIKALQRSAYIREEEIRHLAGRLLKYSVKGNNSKVDMKHMFALLTSSV
IMRVAAGKRHVPVVDEDTKEEKRLLQQFKDLFFPIVSMTICDFIPVMRLIGYKGIEKSMI
QLAEKREVFLNGLVAEVRARRAAAGKTTVGENGEGVSVIDIILNLQETDPEFYTDDIVKG
IVMMMFIAGTETSTATLEWAFTLLMQHPEKMRKLQAEIDSVVGDSRFVCESDFANLPYLR
AVVKETLRMYPPAPLSLPHFSNQACTVGGYDIPKGTMLFANLWTMQRDPKVWDEPDAFKP
ERFEEYDREGGYQYAPFGVGRRSCPGGGMGTHISSLGIGILMQCFDWGHLGMDEDMSHAR
GTASLSRKKPLEADFTPRTKMAGLLSQL

1.3.2. Locus 928

>hpa_locus_928_iso_9_len_1841_ver_2
AATGGAAGAGAGCACGTAGTAGCTAGTCGCTAGTGGTTCTCTCTCCTGAGTGAAGCACGA
GTGATCTAGCTACTACGTGCTCTCTTCCATTTGCTTTTCTCAAAATGGACGGTTTATACT
TAAAACTAGCCCTAACCCTTCTCCTCTCCATTTTTATCGCCTACAAGTTTTTCTTGCGTA
AGAAGCACAACTATCCACCAGGCCCACGTGCCCTCCCCATCATAGGCAACCTCCACCTCA
TCAAGCCCGGCCCACTCTATTTGTACTACGAAACGGTATCGAAGCGCCTCGGCCCGATCA
TTCTCTTCTGGATGGGCAAGGTCCCGGTTATGGTCATCTCCTCCCCGGAGGCCTTTGAGG
AGTGCTTCACCAAGAACGACCTCATCTTCGCGGACCGGCCGAAGACCGTGTCCGGCGACC
ACTTGACCTACAACTACACCTTCATAGTCTTCGCTCCTCACGGCCCCATCTGGAAGAGCC
TCCGCCGGCTTACCTTCTCCGAGATCTTCTCCCAGAAGGCCCTCCAGCGGTCTGCCTTCG
TCCGCGAGGAGGAGGCTCGCCACTTCGTCGGCCGGCTTCTGAAGTGCACCGCCGGCGGGA
ACAAACCCAAGCTAGACATGAAGTACATGCTCTCCCTGCTGACGTCCAGCGTGATCATGA
GGGTGGCAGCTGGAAAACGGCACGTGGCGGTCTCCGACGAGGACACCGAGGAGGAGAAGC
GGCTGATGCAAGAGTTCAAGGACCTCTTCTTCCTTATCGTCTCCATGACGCCTTGTGACT
TCATCCCGGTGCTGAGAATGATCGGTTTCAAGGGCATCGAGAAGAGCATAATCCAGTTGG
GAAAGAAAAGGGAGGTATTCTTGAACAAGTTGCTCGTCGAAGTTAGGGCGCGGAGAAGGG
CGGATGCCGAGAAGCGGTTGAACGGCGAGAAGCGGAACGCGGTTGTCGATACCTTGCTCA
ACTTGCAGGAGACCGATCCTGAGTTTTACACCGATGACATTATCAAGGGAATTATCACGA
TGATGTTTATAGCCGGGACCGAAACGTCAACTGCCACGTTGGAGTGGGCGTTGACCCTCC
TCTTGCAGCACCCAGAGAAGATGGAAAAGCTGCGGGCCGAGATCGACGGGGTCGTGGGAG
ACACTCGGCTCGTGAACGAGTCCGACTTCTCAAAACTTCCGTACCTCAAGGCCGTCGTGA
AGGAGACGATGAGGCTGTTTCCTCCGTCCCCGCTCTTGCTACCTCGTTACTCGAGCGAGG
CTTGCACTATCGGNNNNNNNNNNGCTACTACGTCCCCAAGGGAACCATGCTCCTGGTCAA
CGTATGGACCATGCACAGGGACCCAAAGCTATGGGACGAACCGAACGAGTTCAAGCCCGA
GAGGTTCGTGGCAGGTGACCTTGACGGGTACCTGTACCAGCCGTTTGGCCATGGGAGGAG
GATCTGCCCAGGGGCCGGAATGGGAACCCATATCGTGTGCGAGACATTGGGAATGCTCGT
GCAGNNNNNNNNNNTTTGATTGGGACCAGATCGGTATGGACGAGGACATGACCCACCAAG
CCGGTGCGGCTAGTATGTCCAAGAACAAGCCATTGGAGGCAAGCTTCGCTCCTCGATCCG
AGATGGCCAAGCTCCTTTCCCAACTCTAGCTAGACGAGCCGCGGATCAATTATATGTTTC
TATGTATAAAACTTTGTAATAATGGCATGACTCTCGGACGTGCATGTGTTCTAGTTTGTG
TATTGTGTTTGATGAGGCTATCCCATGTTGATTTAGCTAGATGTGTTCAGTATTTCTGTT
AATTTTGTGTTCATCTTTTCCGCGTTGGAATAATATATTCT

>hpa_locus_928_iso_9_len_1841_ver_2
MDGLYLKLALTLLLSIFIAYKFFLRKKHNYPPGPRALPIIGNLHLIKPGPLYLYYETVSK
RLGPIILFWMGKVPVMVISSPEAFEECFTKNDLIFADRPKTVSGDHLTYNYTFIVFAPHG
PIWKSLRRLTFSEIFSQKALQRSAFVREEEARHFVGRLLKCTAGGNKPKLDMKYMLSLLT
SSVIMRVAAGKRHVAVSDEDTEEEKRLMQEFKDLFFLIVSMTPCDFIPVLRMIGFKGIEK
SIIQLGKKREVFLNKLLVEVRARRRADAEKRLNGEKRNAVVDTLLNLQETDPEFYTDDII
KGIITMMFIAGTETSTATLEWALTLLLQHPEKMEKLRAEIDGVVGDTRLVNESDFSKLPY
LKAVVKETMRLFPPSPLLLPRYSSEACTIGXXXATTSPREPCSWSTYGPCTGTQSYGTNR
TSSSPRGSWQVTLTGTCTSRLAMGGGSAQGPEWEPISCARHWECSCXXXXFDWDQIGMDE
DMTHQAGAASMSKNKPLEASFAPRSEMAKLLSQL

1.3.3. Locus 8128

>hpa_locus_8128_iso_3_len_1785_ver_2
GTCCACACCAACGCATCATAGTTCTTACACTATCCATAACATCGATGGAATTGTATTTGT
ATCTAGCCGCCTTCTTCTTGGCTTTTTCCGCTTTCAAAATCTTGTCTCGAAAAAAGCACA
ATTTTCCGCCTGGCCCACGGACTCTCCCTGTCGTCGGTAATCTCCACCTCATCAAACCAC
CGCTCTACCGTTTTTTCCAGACAGCTTCTGGCAGATTCGGCCCAATCATGTTATTTTGGA
TGGGCCGGACCCCGTTCTTGGTCGTCTCCTCCCCGGCCGCCGTCGAGGAGTGCTTCACTA
AGAACGACATCATCTTCGCCGACCGGCCCAACTCCATGTCCGGCGACCACCTCACTTACA
ACTACNNNNNNNNNNNNTTACAACTACCAGTTCCTCGTCTGGGCCCACCACGGCCCGCTT
TGGAAGAACCTCCGCCGCCTCACCTACACCGAGATCTTCTCCCTCAAAGCCCTCCAGAGG
TCCGCCTTCGTCAGGGAGGAGGAGACCAGCCACTTCGCCGGAAGACTGCTCCGGTACTGC
TCCTCCGGCAAGACGAAGCTGAACATGAAGTACATGTTCTCCCTGCTGGCGTCCAGCGTG
ATCATGAGGGTGGCGGCCGGGAAGCGGCACGTGGCAGTCCACGACGAGGACACGGAGGAA
GAGAGGCGGCGGATCAAGGAGTTCAAGGAGCTGTTCTTCCCCGTCGTGACGCCGGCGGTG
GGAGATTATATTCCGGTGCTGAGGACGATCGGGTACAAGGGGGTCGAGAAGGGGATGATC
CGGCTGGCCCAGAAGAGGGACGTGTACTTGAACAACTTGGTCGCGGAGGTTAGGGCAAAG
AGGGCGGCTTCCGGCAAGAGGTCAAGCGGCGAGGACGGAGACAAGCTCAGCTCCGTCGCT
GACACCGTTCTCAACTTGCAGGAGACCCAACCTGAGGTTTACACCGATGACATCGTCAAG
AGTTTGATCATGATGATGTTTATCGCCGGAACCGAAACGTCAACGACCGTGCTGGAGTGG
TCGTTGACCCTCCTCCTGCAACACCCAAATAAGATGCAAAAGTTACAAGCCGAGATCGAC
GGCATCACCGGAGAGAGCCGCCTGATGAACGAATCCGACCTCCCCAACCTCCCCTACCTC
CGGGCCGTCGTCAAGGAGGCGTTAAGGTTGTTCCCTCCGGTCCCACTCTTGCTGCCTCAC
TTCTCGACCGATGCATGCACCGTTGGAGGCTACGACATCCCCAAAGGAACCATGCTCTTG
GTCAACGTGTGGGCCATGCACAGGGACCCGAAGCTGTGGGACGAACCCACCGAGTTCAGG
CCCGAGAGGTTCGTCGAAACGAACCTCGACACGTACCTGCTCGTGCCGTTTGGTGTTGGG
AGGAGGGCATGTGCTGGAAACACGATGGGAACCCACATGGTGTGCGTGGCGTTGGGCGTG
CTCATGCAGTGCTTCGACTGGGACCAGCTCGGCATGGACGAGGACATGAGCCACGCCGTT
GGCGCGGGTACCTTGTCCAAGGGCAAGCCATTGGAGGCGAGCTTCGCACCTCGGCCCAAG
ATGACCAACCTCCTCAGTTCTCTACTCTCTTAGTCCCAAAGATTGTGTTCGGACCTATTT
TATAAGTTTCTCTTTCAATAAAGCGTCTTTTGTGCGCCTAATAAAAATTTTAATGCTCGG
ATTTGTGTACCGAGTTTAAGTTGCGTGCTTGTACCATTAAATGTAACATTTCGGATAAAA
TCAAATATTTATGTTAGTTAATTATTATCAATAATGCTATTCCCG

>hpa_locus_8128_iso_3_len_1785_ver_2
XPHQRIIVLTLSITSMELYLYLAAFFLAFSAFKILSRKKHNFPPGPRTLPVVGNLHLIKP
PLYRFFQTASGRFGPIMLFWMGRTPFLVVSSPAAVEECFTKNDIIFADRPNSMSGDHLTY
NYXXXXLQLPVPRLGPPRPALEEPPPPHLHRDLLPQSPPEVRLRQGGGDQPLRRKTAPVL
LLRQDEAEHEVHVLPAGVQRDHEGGGREAARGSPRRGHGGREAADQGVQGAVLPRRDAGG
GRLYSGAEDDRVQGGREGDDPAGPEEGRVLEQLGAEVRAKRAASGKRSSGEDGDKLSSVA
DTVLNLQETQPEVYTDDIVKSLIMMMFIAGTETSTTVLEWSLTLLLQHPNKMQKLQAEID
GITGESRLMNESDLPNLPYLRAVVKEALRLFPPVPLLLPHFSTDACTVGGYDIPKGTMLL
VNVWAMHRDPKLWDEPTEFRPERFVETNLDTYLLVPFGVGRRACAGNTMGTHMVCVALGV
LMQCFDWDQLGMDEDMSHAVGAGTLSKGKPLEASFAPRPKMTNLLSSLLS

1.3.4. Locus 51544

>hpa_locus_51544_iso_1_len_1149_ver_2
TTTCTTTCTTAATTCCCATTAGCTCTTCACAAAATGGAGGACTTGTACTTCTACCTCGCC
GTTATTCTATCAGCTTTTGTCGCTTACAAGGTGTTGTCCAGAAAGAAACATAACTACCCT
CCAAGCCCAAGGGCGCTCCCCATCTTTGGCCACCTCCACCTCGTCAAATACCCTCTCTAT
ATGTACTTCCAAACCGTCTCCAACCTCCTCGGCCCGATCCTGCTCTTCTGGATGGGCAAG
GTCCCGGTCCTGGTCATCTCCTCTCCCGCGGCCTTCGAGGAGTGCTTCACCACGAACGAC
GTTGTATTCGCCGACAGGCCTAAGTCCATCACCACTGACCACCTAACCTACGACTACACC
TTCATGGTCTGGGCCCCACATGGCCAGCTCTGGAAAAGCCTCCGCCGCCTCACATTCTCC
GAGATCTTCTCCCAGAGAGCCCTCCAGAGAACCGCCATCGTTAGGGAAGAGGAGATCCGC
CACTTTGTCGGCCGGCTCCTTCAATACACTGTAGGCGGGGGCAAGCCCAAGGTTGACATG
CAGCGCATGTTGTCTCTGTTAGCGTCCAACGTGATCATGAAAGTTGCGGCAGGGAAGAGA
CACGTGCCGATCAGTGACGAGGACACCGACGAGGAGGAGCGGCTGCTCCAGCAGTTCAAG
GACCTCTTCTTTGCTATTGTGGCGATGACGCCATGCGACTTCATCCCATTTCTGAGGTTT
ATCGGGTACAAGGGGACAGAGAAGAGAATCATCCGACTGGGGGAGAAGAGGGATGTCTTC
CTTAACGGCTTACTTGCCAATATTAGGGCAAAGAGGGCGGCTGCTCGGAAGGGGACAAAT
AATAAGGAGGGCGACAAGCTTAGCTCAGTGGTGGATACCATCCTTGACTTGCAGGAGACA
GACCCTCTGTTCTTTACCGACGAAATCGTCAAGGGCATTACAATGATGATGTTCAAGATC
GGAAGAGCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNTCCTTATGCAGCACCCGGAGAAGATCCAAAAATTACAGGCCGAAATTGACAACGTC
GTAGGAGATAGAAGATTCCTCAATGAATATGACTTGGCCAACCTTCCTTACCTTAGAGCA
GTCGTGAAG

>hpa_locus_51544_iso_1_len_1149_ver_2
MEDLYFYLAVILSAFVAYKVLSRKKHNYPPSPRALPIFGHLHLVKYPLYMYFQTVSNLLG
PILLFWMGKVPVLVISSPAAFEECFTTNDVVFADRPKSITTDHLTYDYTFMVWAPHGQLW
KSLRRLTFSEIFSQRALQRTAIVREEEIRHFVGRLLQYTVGGGKPKVDMQRMLSLLASNV
IMKVAAGKRHVPISDEDTDEEERLLQQFKDLFFAIVAMTPCDFIPFLRFIGYKGTEKRII
RLGEKRDVFLNGLLANIRAKRAAARKGTNNKEGDKLSSVVDTILDLQETDPLFFTDEIVK
GITMMMFKIGRAXXXXXXXXXXXXXXXXXXXLMQHPEKIQKLQAEIDNVVGDRRFLNEYD
LANLPYLRAVVK

1.4. Sequences of the cloned genes

1.4.1. HcCYP81AA1 (B3′H/1,3,7-TXS); NCBI accession number: KT716863

>*HcCYP81AA1*

GACAGCCCACAACTCTCAAAATACACATCACATAAAAATAAGTCGTCGTCTTGTAAAACAGAAGAGAGAG
ATCAACTCCTTTTAAAAAA**ATG**GAGGACTTGTACTTGTACCTCGCCGTTCTCGTCTCCGTCGTCGTCGGC
TACAAGGTTCTCTTCGGTAAGAAGCTCAAATACCCGCCGGGCCCACGTGCACTTCCCATCGTCGGTCACC
TCCCCGTCCTCAAGCCCCCGCTCTACCAGTACTTCGAGAGCGTCTCCAAGAAACTTGGCCCGGTCTTGCT
CTTCTGGATGGGCAAGGTCCCCGTCCTCGTTGTCTCCTCCCCGGAGGCCTTTGAGGAGTGCTTCACCAAG
AACGACGTCGTCTTTGCCGACCGCCCCAAGTCCATGTCCGGCGAGCATCTCACCTATGACTACACCTTCC
TCGTGTGGGCCCCCCACGGCCCCATCTGGAAGAGCCTCCGTCGGCTCACCTTCTCGGAGATCTTCTCCAT
CAAGGCCCTCCAGAGGTCGGCCTTCATCCGGGAGGAGGAGATCCGGCACCTCGCCGGCCGCCTGCGGAAG
TACACCGTCAACGGGAACAACTCCAAGGTGGACATGAAGCACTTGTTCGCCCTGCTGACGTCGAGCGTGA
TCATGCGGGTGGCGGCCGGGAAGAGGCACGTGCCGGTCGCCGACGAGGACACCAAGGAGGAGAAGCGGCT
GCTCCAGCAGTTCAAGGACCTCTTCTTCCCTATCGTCTCGATGACCGTATGTGACTTCATCCCGGTGATG
AGGCTGATCGGGTACAAGGGGATCGAGAAGAGCATGATCCAGCTGGCCGAGAAGAGGGAGGTCTTCCTTA
ACGCCCTGGTCGCCGAGGTCAGAGCCAGGAGGGCCGCCGGCGGGAAGACGACAGTTGGCGAGAACGGCGA
AGGGGTCTCTGTGATCGATATCATCCTCAACTTGCAGGAGACAGACCCCGAGTTTTACACCGATGCCATC
GTCAAGGGTATTGTTATGATGATGTTTATCGCCGGGACCGAGACGTCGACCGCCACGTTGGAGTGGGGGT
TCACTCTCCTAATGCAACACCCAGAGAAGATGCGCAAACTACAAGCCGAGATCGACAGCGTCGTAGGAGA
CAGCCGCTTCGTGAGCGAAGCCGACTTCGCCAATCTTCCCTACCTAAGGGCCGTCGTGAAGGAGACACTA
AGGCTGTACCCGCCGGCCCCACTCTCCCTCCCCCACTTCTCTAACCAAGCCTGCACCGTCGGAGGCTACG
ACATTCCCAATGGAACCATGCTCTTCGCCAACTTGTGGGCAATGCAGAGAGACCCTAAAGTATGGGACGA
GCCTGACGCGTTCAAGCCCGAGAGGTTCGAGGAATACGACCGTGAAGGCGGGTACCAGTACGCGCCGTTT
GGTGTCGGGAGGAGGTCGTGCCCCGGCGGAGGGATGGGGACCCACATTTCGTGCCTGGGGATTGGAATCC
TGATGCAGTGCTTCGACTGGGGCCAGCTCGGTATGGACGAGGACATGAGCCACGCGCGTGGTACCGCAAG
TTTGTCGAGGAAGAAGCCGTTGGAGGCTGACTTCACGCCTCGCACCAAGATGGCCGGTCTCCTTTCCCAG
CTC**TAA**ACCTCTCTCGCGCCGGCTTATTTGTTCAATTTTATTTTTACTTTTTCTGAAACTATTGAATAAT
GCTGATATGTACCTCGAGACGTGGTTCAAACTTCAAATCTTTTGGACTGTATACTTTTTTTTTATATAAA
TGTGTACTTGTTGGAAAAATAATTGTAGACCTGTGTACCGAATTTTAGTATAAAAAAAAAAAAAAAAAA
AAAAAAAAAA

>HcCYP81AA1

MEDLYLYLAVLVSVVVGYKVLFGKKLKYPPGPRALPIVGHLPVLKPPLYQYFESVSKKLGPVLLFWMGKV
PVLVVSSPEAFEECFTKNDVVFADRPKSMSGEHLTYDYTFLVWAPHGPIWKSLRRLTFSEIFSIKALQRS
AFIREEEIRHLAGRLRKYTVNGNNSKVDMKHLFALLTSSVIMRVAAGKRHVPVADEDTKEEKRLLQQFKD
LFFPIVSMTVCDFIPVMRLIGYKGIEKSMIQLAEKREVFLNALVAEVRARRAAGGKTTVGENGEGVSVID
IILNLQETDPEFYTDAIVKGIVMMMFIAGTETSTATLEWGFTLLMQHPEKMRKLQAEIDSVVGDSRFVSE
ADFANLPYLRAVVKETLRLYPPAPLSLPHFSNQACTVGGYDIPNGTMLFANLWAMQRDPKVWDEPDAFKP
ERFEEYDREGGYQYAPFGVGRRSCPGGGMGTHISCLGIGILMQCFDWGQLGMDEDMSHARGTASLSRKKP
LEADFTPRTKMAGLLSQL

1.4.2. HcCPR2 (NADPH-dependent cytochrome P450 oxidoreductase); NCBI accession number: KT777456

>HcCPR2

GGCCAAAAAAAAAAACAAAACTCTCTTCTCCCTCCCACCAACCAGCCTCGAAACGCACCGTTTTAGAGAG
AGAAAGAGAGGGAGAAACAGAACTCGCGCGAAGTAACCGCTAAACTA**ATG**GAACCGACGGGGAGCTCGAC
GGCGATGAAGGTGTCGCCGCTTGATCTAATGGCTGCGATTTTAAAGGGCGGCCGTGTGGACCCCGCGAAC
GCCTCGTCCGCCGGTCCCGAGGTGGTTTCGATGATACTGGGGAACAGGGAGACGGTTGTTATGGTCGTGG
CCACCTCCATCGCCGTGATCATCGGATGCGCGGTCTTCCTCTTCTGGCGGCGGTCGTCGTCGTCCCGGGT
TAAGCCGAAGCCCCACCCGCTCCGGCCGATCGCTATGAGAGATCTCGACCTCGACGAGGCCGTCGACGGG
AAGAAGAAGGTCACTATCTTGTTCGGCACCCAAACCGGAACCGCCGAAGGCTTTGCTAAGGCTTTGGCCG
AGGAAGCAAAGGCGCGGTATGACAAGGCGGTCTTCAAAGTTGTTGACTTGGATGACTATGCTGCGGATGA
TGAAGAGTATGAGGAGAAGTTGAAGAAGGAAGATTTGGTTGTCCTCTTCTTGGCCACGTACGGAGACGGT
GAGCCGACAGATAATGCTGCTAGGTTTTATAAATGGCTTACCGAGGGGAATGAGAGAAGTGAGTGGCTGA
AACACCTCCAGTTTGCAGTGTTTGGCTTAGGGAACAGGCAGTATGAGCACTTTAACAAGGTTGCAATTGT
GGTGGATGAGATTCTTGCCAAAGAAGGTGCAAAGCGCCTTGCTCCAGTCGGTCTCGGAGATGATGATCAA
TGCATCGAAGATGATTTCAATGCATGGCGAGAGGCATTGTGGCCCGAAATGGATCAGTTGCTCCACGATG
AAGATGATGTGACAACTGTTGCTACTCCATACACTGCTGCAGTTCCTGAATATCGTGTAGTGTTTAATGA
CTCTGTTGATCCAAATGTTGAGGATAAAAACTGGAGCAATGCAAATGGCCATGCTGTTTATGATGCTCAA
CATCCATGCCGGGCAAATGTTGCAGTCAGGAAGGAGCTTCATTCTCCTGCATCTGATCGTTCTTGCACCC
ATCTAGAGTTTGACATAGCTGGCACTGGACTTTCTTATGAAACCGGGGACCATGTTGGTGTTTATTGCGA
GAACTTAGAGGAAACAGTGGAAGAAGCCTTGCAATTATTGGGCTTGTCGCCAGATACTTACTTTTCTCTT
CATAGCAATAAAGAGGACGGGACACCCCTTGGAGGTTCTCTGTCACCTCCTTTCCCACCTTGCACGTTAA
GAACAGCTCTGGCTAGATATGCTGACCTTTTGAGTTCTCCCAAAAAGTCTGCACTGCTGGCTTTAGCTGC
ACATGCCACTGATCCAACTGAAGTTGATCGACTGAGGCATCTTGCATCCCTGGCGGGAAAGGATGATTAT
GCACAATGGATGGTTGCCAATCAAAGAAGCCTCCTTGAGGTCATGGCTGAATTTCCTTCTGCCAAGCCCC
CTCTTGGTGTCTTTTTTGCTGCAGTTGCCCCACGCTTACAGCCAAGATACTATTCTATCTCGTCATCACC
CAAGATTGCACCAACTAGGATTCATGTGACCTGTGCCCTAGTTCTCGAGAAGACACCAACCGGTCGGATT
CATAAGGGAATTTGCTCAACTTGGATGAAGAATGCCATCCCCATGGAGAAAAGCAATGATTGCAGCTGGG
CTCCAGTTTTTGTTAGACAATCAAACTTTAAACTTCCAGCAGATCCTAAGGTCCCTGTTATAATGATTGG
TCCTGGTACTGGCTTGGCTCCATTCAGGGGTTTCCTGCAGGAAAGGTTAGCCCTGAAAGAATCCGGAGCA
GAACTGGGACCAGCCGTTTTGTTTTTCGGCTGCAGAAATCGCAGGATGGATTTCATCTACGAAGATGAAC
TAAACAATTTCCTTGAATGTGGCGCACTTTCGGAGCTGGTTATCGCCTTCTCCCGTGAAGGACCCACCAA
GGAGTACGTGCAACATAAAATGGCCGAGAAGGCCTCGGATATTTGGAACATGATTTCTCAAGGAGGTTAT
CTTTACGTTTGTGGGGATGCCAAAGGAATGGCAAAAGATGTGCACAGAACTCTCCACACCATTGTCCAAG
AACAGGGATCACTTGATAGCTCAAAAACAGAGAGCATGGTCAAGAATTTACAAATGAATGGGCGGTACCT
TCGCGACGTATGG**TGA**TGAGTTAAAAGAGCCAAATTCCTACATTTACATGTAGTGGGACCACCCCAACCG
TTGGATGATGCTGCGAAATTAATTTCGGAAAGGCTACCGCTACATTTTTCTTCTATGGTTGCTTTATTCT
TTTTCTTCTTATTTGGAAGATAAAATTATTCCTGTAATCTATAAATGACTCACAATAGCGCCATAAATGG
TATTAGGTGTCTTAGTTATTAAATATATATATCTTACTGTATTTCTATGTAAACCGAAATATTAAATATT
ATCATGTAGTGTACTACTTCTCTATCCATTCACTCGGCGCAGTGGAAAAATAAAAAAAAAAAAAAAAAAA
AAAAAAAAA

>HcCPR2

MEPTGSSTAMKVSPLDLMAAILKGGRVDPANASSAGPEVVSMILGNRETVVMVVATSIAVIIGCAVFLFW
RRSSSSRVKPKPHPLRPIAMRDLDLDEAVDGKKKVTILFGTQTGTAEGFAKALAEEAKARYDKAVFKVVD
LDDYAADDEEYEEKLKKEDLVVLFLATYGDGEPTDNAARFYKWLTEGNERSEWLKHLQFAVFGLGNRQYE
HFNKVAIVVDEILAKEGAKRLAPVGLGDDDQCIEDDFNAWREALWPEMDQLLHDEDDVTTVATPYTAAVP
EYRVVFNDSVDPNVEDKNWSNANGHAVYDAQHPCRANVAVRKELHSPASDRSCTHLEFDIAGTGLSYETG
DHVGVYCENLEETVEEALQLLGLSPDTYFSLHSNKEDGTPLGGSLSPPFPPCTLRTALARYADLLSSPKK
SALLALAAHATDPTEVDRLRHLASLAGKDDYAQWMVANQRSLLEVMAEFPSAKPPLGVFFAAVAPRLQPR
YYSISSSPKIAPTRIHVTCALVLEKTPTGRIHKGICSTWMKNAIPMEKSNDCSWAPVFVRQSNFKLPADP
KVPVIMIGPGTGLAPFRGFLQERLALKESGAELGPAVLFFGCRNRRMDFIYEDELNNFLECGALSELVIA
FSREGPTKEYVQHKMAEKASDIWNMISQGGYLYVCGDAKGMAKDVHRTLHTIVQEQGSLDSSKTESMVKN
LQMNGRYLRDVW

1.4.3. HpCYP81AA1 (B3′H/1,3,7-TXS); NCBI accession number: KT716864

>*HpCYP81AA1*

ATGGAGGACTTGTATCTGTACGTCGCCGTTCTTGTCTCCGTCATCATCGGCTACAAGGTTCTCTTCGGTA
AGAAACTCAACTACCCGCCGGGCCCACGGGCATATCCCATCGTCGGACATCTCCCCGTCCTCAAGCCCCC
GCTCTACCAGTACTTCGAGAGCGTCTCCAAGAAGCTCGGCCCAGTCCTTCTCTTCTGGATGGGCAAGGTC
CCCGTCCTCGTCGTCTCCTCCCCGGAGGCCTTCGAGGAGTGCTTCACAAAGAACGACGTCGTCTTCGCCG
ACCGCCCGAAGTCCATGTCAGGTGAGCACCTCACCTACGACTACACCTTCCTCGTGTGGGCCCCGCACGG
CCCCATCTGGAAGAGCCTCCGACGTCTCACCTTCTCCGAGATCTTCTCCATCAAGGCCCTCCAGAGGTCC
GCCTACATCCGGGAGGAGGAGATCCGCCACCTGGCCGGCCGCCTTCTCAAGTACTCCGTCAAAGGGAACA
ACTCCAAGGTGGACATGAAGCACATGTTCGCCCTGCTGACGTCAAGCGTGATCATGCGGGTGGCGGCCGG
GAAGAGGCACGTGCCGGTCGTCGACGAGGACACCAAGGAGGAGAAGCGGCTGCTCCAGCAGTTCAAGGAT
CTCTTCTTTCCTATCGTCTCCATGACCATATGCGACTTCATCCCGGTGATGAGGCTGATCGGTTACAAGG
GGATCGAGAAGAGCATGATCCAGCTGGCCGAGAAGAGGGAGGTCTTCCTCAACGGCCTTGTCGCCGAGGT
CAGAGCCAGGAGGGCGGCCGCCGGGAAGACCACTGTTGGCGAGAACGGCGAAGGGGTCTCCGTTATCGAT
ATCATCCTCAACTTGCAAGAGACAGACCCTGAGTTTTACACGGATGACATCGTCAAGGGAATTGTTATGA
TGATGTTTATTGCCGGGACCGAGACGTCGACCGCCACGTTGGAGTGGGCGTTCACACTGCTGATGCAACA
CCCTGAGAAGATGCGGAAACTACAAGCCGAGATCGACAGCGTCGTCGGAGACAGCCGCTTCGTGTGCGAA
TCCGACTTCGCCAACCTTCCCTACCTACGGGCGGTCGTGAAGGAGACGCTCAGGATGTACCCGCCGGCCC
CACTCTCCCTCCCCCACTTCTCCAACCAAGCCTGCACCGTCGGAGGCTACGACATTCCCAAGGGAACCAT
GCTCTTCGCCAACCTGTGGACAATGCAGAGAGACCCCAAAGTATGGGACGAGCCCGACGCGTTCAAGCCC
GAGAGGTTCGAGGAGTACGACCGCGAAGGAGGGTACCAGTACGCGCCGTTCGGTGTCGGGAGGAGGTCTT
GCCCCGGCGGAGGGATGGGGACCCACATTTCGTCCCTGGGGATTGGAATTCTTATGCAGTGCTTCGACTG
GGGCCACCTCGGCATGGACGAGGACATGAGCCACGCGCGTGGTACCGCGAGTCTGTCAAGGAAGAAGCCA
TTGGAGGCGGACTTCACGCCTCGCACCAAGATGGCCGGTCTCCTCTCCCAGCTC**TAA**

>HpCYP81AA1

MEDLYLYVAVLVSVIIGYKVLFGKKLNYPPGPRAYPIVGHLPVLKPPLYQYFESVSKKLGPVLLFWMGKV
PVLVVSSPEAFEECFTKNDVVFADRPKSMSGEHLTYDYTFLVWAPHGPIWKSLRRLTFSEIFSIKALQRS
AYIREEEIRHLAGRLLKYSVKGNNSKVDMKHMFALLTSSVIMRVAAGKRHVPVVDEDTKEEKRLLQQFKD
LFFPIVSMTICDFIPVMRLIGYKGIEKSMIQLAEKREVFLNGLVAEVRARRAAAGKTTVGENGEGVSVID
IILNLQETDPEFYTDDIVKGIVMMMFIAGTETSTATLEWAFTLLMQHPEKMRKLQAEIDSVVGDSRFVCE
SDFANLPYLRAVVKETLRMYPPAPLSLPHFSNQACTVGGYDIPKGTMLFANLWTMQRDPKVWDEPDAFKP
ERFEEYDREGGYQYAPFGVGRRSCPGGGMGTHISSLGIGILMQCFDWGHLGMDEDMSHARGTASLSRKKP
LEADFTPRTKMAGLLSQL

1.4.4. HpCYP81AA2 (B3′H/1,3,5-TXS); NCBI accession number: KT716865

>HpCYP81AA2

ATGGACGGTTTATACTTAAAACTAGCCCTAACCCTTCTCCTCTCCATTTTTATCGCCTACAAGTTTTTCT
TGCGTAAGAAGCACAACTATCCACCAGGCCCACGTGCCCTCCCCATCATAGGCAACCTCCACCTCATCAA
GCCCGGCCCACTCTATTTGTACTACGAAACGGTATCGAAGCGCCTCGGCCCGATCATTCTCTTCTGGATG
GGCAAGGTCCCGGTTATGGTCATCTCCTCCCCGGAGGCCTTTGAGGAGTGCTTCACCAAGAACGACCTCA
TCTTCGCGGACCGGCCGAAGACCGTGTCCGGCGACCACTTGACCTACAACTACACCTTCATAGTCTTCGC
TCCTCACGGCCCCATCTGGAAGAGCCTCCGCCGGCTTACCTTCTCCGAGATCTTCTCCCAGAAGGCCCTC
CAGCGGTCCGCCTTCGTCCGCGAGGAGGAGGCTCGCCACTTCGTCGGCCGGCTTCTGAAGTGCACCGCCG
GCGGGAACAAACCCAAGCTAGACATGAAGTACATGCTCTCCCTGCTGACGTCCAGCGTGATCATGAGGGT
GGCAGCTGGAAAACGGCACGTGGCGGTCTCCGATGAGGACACCGAGGAGGAGAAGCGGCTGATGCAAGAG
TTCAAGGACCTCTTCTTCCCTATCGTCTCGATGACGCCGTGTGACTTCATCCCGGTGCTGAGGATGATCG
GGTTCAAGGGCATCGAGAAGAGCATAATCCAGTTGGGAAAGAAGAGAGAGGTATTCTTGAACAAGTTGCT
CGTCGAAGTTAGGGCGAGGAGAGGGGCGGATGCCGAGAAGCGGTTGAACGGCGAGAAGCCGAACGCGGTG
GTCGATACCTTGCTCAACTTGCAGGAGACCGATCCTGAGTTTTACACCGATGACATTATCAAGGGAATTA
TCACGATGATGTTCATAGCCGGGACCGAAACGTCAACTGCCACGTTGGAGTGGGCGTTGACCCTCCTCTT
GCAGCACCCAAAGAAGATGGAAAAGCTGCGGGCCGAGATCGACGGGGTCGTGGGAGACACTCGGCTCGTG
AACGAGTCCGACTTCTCAAAACTTCCGTACCTCAAGGCCGTCGTGAAGGAGACGATGAGGCTGTATCCTC
CGTCCCCGCTCTTGCTACCTCGTTATTCGAGCGAGGCTTGCACTATCGGAGGCTACTACGTCCCCAAGGG
AACCATGCTCCTGGTCAACGTATGGACCATGCACAGGGACCCAAAACTATGGGACGAACCGAACGAGTTC
AAGCCCGAGAGGTTCGTGGAAGGCGACCTTGAGGGGTACCTGTACCAGCCGTTTGGCCACGGGAGGAGGA
TCTGCCCAGGAGCCGGAATGGGAACCCATATCGTCTGCGAGACATTGGGAATGCTCGTGCAGTGCTTTGA
CTGGGACCAGATCGGTATGGACGAGGACATGACCCACCAAGCCGGTGCGGCTAGTATGTCCAAGAACAAG
CCATTGGAGGCAAGCTTCGCACCTCGATCCGAGATGGCCAAGCTCCTTTCCCAACTC**TAG**

>HpCYP81AA2

MDGLYLKLALTLLLSIFIAYKFFLRKKHNYPPGPRALPIIGNLHLIKPGPLYLYYETVSKRLGPIILFWM
GKVPVMVISSPEAFEECFTKNDLIFADRPKTVSGDHLTYNYTFIVFAPHGPIWKSLRRLTFSEIFSQKAL
QRSAFVREEEARHFVGRLLKCTAGGNKPKLDMKYMLSLLTSSVIMRVAAGKRHVAVSDEDTEEEKRLMQE
FKDLFFPIVSMTPCDFIPVLRMIGFKGIEKSIIQLGKKREVFLNKLLVEVRARRGADAEKRLNGEKPNAV
VDTLLNLQETDPEFYTDDIIKGIITMMFIAGTETSTATLEWALTLLLQHPKKMEKLRAEIDGVVGDTRLV
NESDFSKLPYLKAVVKETMRLYPPSPLLLPRYSSEACTIGGYYVPKGTMLLVNVWTMHRDPKLWDEPNEF
KPERFVEGDLEGYLYQPFGHGRRICPGAGMGTHIVCETLGMLVQCFDWDQIGMDEDMTHQAGAASMSKNK
PLEASFAPRSEMAKLLSQL

1.4.5. HpCYP81AA3; NCBI accession number: KT716866

>*HpCYP81AA3*

ATGGAATTGTATTTGTATCTAGCCGCCTTCTTCTTGGCTTTTTCCGCTTTCAAACTCTTGTCTCGAAAGA
AGCACAATTATCCACCTGGCCCACGGGCTCTCCCGGTCGTCGGTAATCTCCACCTTATCAAACCACCGCT
CTACCGTTTTTTCCAGACAGCTTCTGGCAGATTCGGCCCAATCATGTTATTTTGGATGGGCCGGACCCCG
TTCTTGGTCGTCTCCTCCCCGGACGCCGTCGAGGAGTGCTTCACTAAGAACGACATCATCTTCGCCGACC
GGCCCAACTCCATGTCCGGCGACCACCTCACTTACAACTACCAGTTCCTCGTCTGGGCCCACCACGGCCC
GCTTTGGAAGAACCTCCGCCGCCTCACCTACACCGAGATCTTCTCCCTCAAAGCCCTCCAGAGGTCTGCC
TTCGTCAGGGAGGAGGAGACCAGCCACTTCGCCGGAAGACTGCTCCGGTACTGCTCCTCCGGCAAGACGA
AGCTGAACATGAAGTACATGTTCTCCCTGCTGGCTTCCAGCGTGATCATGAGGGTGGCGGCCGGGAAGCG
GCACGTGGCAGTCCACGACGAGGACACGGAGGAAGAGAGGCGGCGGATCAAGGAGTTCAAGGAGCTGTTC
TTCCCCGTCGTGACGCCGGCGGTGGGAGATTATATTCCGGTGCTGAGGACGATCGGGTACAAGGGGGTCG
AGAAGGGGATGATCCGGCTGGCCCAGAAGAGGGACGTGTACTTGAACAACTTGGTCGCGGAGGTTAGAGC
AAAGAGGAAGGCTTCCGGCAAGAGGTCAAGCGGCGAGGACGGAGACAAGCTCAGCTCCGTCGCTGATACC
GTTCTCAACTTGCAGGAGACCCAACCTGAGCTTTACACCGATGACATCGTCAAGAGTTTGGTCATGATGA
TGTTTATCGCCGGAACCGAAACGTCAACGACCGTGCTGGAGTGGTCGTTGACCCTCCTCTTGCAACACCC
AAATAAGATGCAAAAGTTACAAGCCGAGATCGACGGCATCACCGGAGAGAGCCGCCTCATGAACGAATCC
GACCTCCCCAACCTCCCCTACCTCCGGGCCGTCGTCAAGGAGGCGTTAAGGTTGTTCCCTCCGGTCCCAC
TCTTGCTGCCTCACTTCTCGACCGATGCCTGCACCGTTGGAGGCTACGACATTCCCAAAGGAACCATGCT
CTTGGTCAACGTGTGGGCCATGCACAGGGACCCGAAGCTGTGGGACGAACCCACCGAGTTCAGGCCCGAG
AGGTTCGTCGAAACGAACCTCGACACGTACCTGCTCGTGCCGTTTGGTGTTGGGAGGAGGGCATGTGCTG
GAAACACGATGGGAACCCACATGGTGTGCGTGGCGTTGGGCGTGCTCATGCAGTGCTTCGACTGGGATCA
GCTCGGCATGGACGAGGACATGAGCCACGCGGTTGGCGCGGGTACCTTGTCCAAGGGCAAGCCATTGGAG
GCGAGCTTCGCACCTCGGCCCAAGATGACCAACATCCTCAGTTCTCTACTCTCT**TAG**

>HpCYP81AA3

MELYLYLAAFFLAFSAFKLLSRKKHNYPPGPRALPVVGNLHLIKPPLYRFFQTASGRFGPIMLFWMGRTP
FLVVSSPDAVEECFTKNDIIFADRPNSMSGDHLTYNYQFLVWAHHGPLWKNLRRLTYTEIFSLKALQRSA
FVREEETSHFAGRLLRYCSSGKTKLNMKYMFSLLASSVIMRVAAGKRHVAVHDEDTEEERRRIKEFKELF
FPVVTPAVGDYIPVLRTIGYKGVEKGMIRLAQKRDVYLNNLVAEVRAKRKASGKRSSGEDGDKLSSVADT
VLNLQETQPELYTDDIVKSLVMMMFIAGTETSTTVLEWSLTLLLQHPNKMQKLQAEIDGITGESRLMNES
DLPNLPYLRAVVKEALRLFPPVPLLLPHFSTDACTVGGYDIPKGTMLLVNVWAMHRDPKLWDEPTEFRPE
RFVETNLDTYLLVPFGVGRRACAGNTMGTHMVCVALGVLMQCFDWDQLGMDEDMSHAVGAGTLSKGKPLE
ASFAPRPKMTNILSSLLS